

The case for helping Mexico — now

The money crisis in Mexico is a great setback for civility in the Western Hemisphere and for hopes of greater US non-colonial influence in Latin America. The US Congress must first act on the loan guarantees asked of it.

LESS than a year ago, Mexico seemed on the brink of joining the rich countries of the world. At the time of this journal's survey of "Science in Mexico" (*Nature* 368, 789–804; 1994), there were several cheering tendencies to describe. The North American Free Trade Agreement (NAFTA) had been signed, and offered even greater benefits to Mexico than to its partners, the United States and Canada. Many industrial companies had independently discovered how to make money by exporting what they manufactured, and seemed destined to get on even faster. Science and technology had been given an honourable place near the centre of public affairs, and the means of awarding research funds (still too meagre) had been made fair, even imaginative. So why has Mexico seemingly collapsed, in just a few weeks, to the point at which bankers are worrying that it will have defaulted on its foreign debts by about now?

To be sure, there were clouds on the horizon last spring. Earlier in the year, there has been a rebellion against the civil governor of the southern state of Chiapas by Mexicans of indigenous descent, only temporarily settled and only then at some political cost to the government. More seriously, the then-impending election (held on 15 August), while advertised (by the Institutional Revolutionary Party, or PRI) to be less undemocratic than its predecessors, left much to be desired. But against the optimism endangered by the government's declared commitment to reform, these clouds seemed no bigger than a man's hand.

What has gone wrong is a simple miscalculation in economics. A year ago, Mexico was running its economy at a deficit with the rest of the world amounting to about 15 per cent of its gross domestic product, but was able to pay its bills because the inward investment of hard currency in its domestic enterprises roughly matched the deficit. At the time, nobody seemed particularly concerned. NAFTA could surely only increase the inward flow of hard currency? But, in the event, the inward flow actually diminished, perhaps because of the diversion of investment funds to other places, perhaps because of economic recovery in the United States. So Mexico's external spending was no longer covered by investment funds, the peso started to slide, investors lost confidence in their investments, the peso slid further ... and so on.

At the beginning of this week, Mexico's official holdings of foreign currency were said to be worth just US\$2 billion. Last week, the central bank even failed to sell the planned instalments of domestic debt that are denominated in US

dollars so as to make them palatable, in uncertain times, to Mexicans. The calculation now is that Mexico needs \$40 billion to keep it afloat, and there is a provisional deal whereby the United States would provide loan guarantees for that amount. (Mexico would be required to repay the United States in oil if the guarantees were ever called in.) But the US Congress, in its new colours, may well decline to sanction the deal. That, of course, would kill off NAFTA. It may do much worse.

The United States has no formal obligation to help out, but every self-interested reason for doing so. Mexico is not merely its nearest neighbour to the south, but also one with which it is inextricably intertwined. The most conspicuous ties may be those of illegal immigration into Texas and California, but the cultural ties go back further, and are more permanent. That the United States is the chief provider of overseas PhD certificates to Mexicans is only one proof of the connection. But the opportunities that will be sacrificed if there is no loan guarantee put the past beyond the pale. NAFTA offered the United States the chance of a mutually beneficial relationship with a latin country to the south that would be so transparently free of the taint of colonialism that others would wish to follow. President Bill Clinton eloquently spelled out those hopes at the Latin-American summit meeting just two months ago. Everybody will be worse off if they come to nothing. □

Per ardua ad astra

US astronomy faces impossible choices unless there is a little more to spend.

ASTRONOMY in the United States is in a bind. Resources are only meagre, and ill-matched with people in post. In an equitable world, US astronomers would draw some comfort from knowing that the problems are similar in Europe. But the world is not an equitable place. And the United States supports an even greater proportion of top-class research in astronomy and astrophysics than in, say, low-temperature physics. Worse, it is only four years since the National Academy of Sciences commissioned (at government behest) John Bahcall of Princeton to say what the next big spending should be. Now the National Science Foundation (NSF), which pays the bills, has put a damper on the proceedings with a report by a panel under Richard McCray (from the



University of Colorado) whose task has been to help the foundation manage the parsimonious years ahead.

McCray and his colleagues have been thinking the unthinkable with a vengeance. If there is no overall increase in the astronomy budget, they say, it will be necessary to close most of the telescopes at Kitt Peak National Observatory in Arizona. That is the unthinkable. McCray also entertains the notion (part of his remit) that there may be either a minimal or a modest increase of the budget of the National Optical Astronomy Observatories (NOAO), which operate Kitt Peak and the Cerro Tololo Interamerican Observatory in Chile. NOAO's difficulty is its commitment to the Gemini project, two new 8-metre telescopes in Hawaii and Chile now being built, which will cost about \$8 million a year to operate. With a minimal increase, other facilities will have to be closed. With a modest increase, NOAO may be able to scrape by.

Closing a substantial part of Kitt Peak is unthinkable because the US national astronomy observatories are a cultural treasure, and not just for the United States. They allow access to world-class telescopes solely on the basis of scientific merit, without regard to the passports of those who use them. McCray wants to keep that principle, but his panel also rightly asks that modern facilities should have precedence over older facilities. That is why Kitt Peak, where most of the telescopes are more than 20 years old, is vulnerable. But apart from its international connections, half of US astronomers depend on Kitt Peak for optical observations in the Northern Hemisphere.

The McCray panel has an ingenious idea for softening that blow: why not make a deal with the private observatories that now control the bulk of the collecting area in the United States? (The 10-metre Keck telescope comprises a large fraction of that total.) The proposal is that they should sell some of their observing time to NSF, spending what they earn on instrumentation, perhaps at the rate of \$4.5 million a year. That could make up for the well-known unwillingness of private benefactors to spend funds on instruments rather than on telescopes. But successful deal-making would require growth in the NSF budget not yet assured.

Meanwhile, NSF has decided that there is no point in building major new scientific facilities without knowing where the operating funds will come from. Future projects, such as the millimetre-wave array (recommended by Bahcall) will be approved only when operating funds have been identified. CERN, the high-energy physics laboratory at Geneva, has always budgeted in that way, with admirable success. If that means fewer new projects in the future, the price may well be worth the benefit of letting the community know where it stands.

Early in this year's US budget cycle, nobody can yet tell which of McCray's assumptions (if any) will apply to NSF's planning for the financial year beginning on 1 October, but the new Congress is making ominous noises. Its members should therefore acknowledge two features of US public support for the exploration of the Universe. First, the national observatories have been among the principal sources of the wealth of new knowledge gathered in recent decades, to the delight of taxpayers in the United States and else-

where. Second, McCray's careful introspection about NSF's financial dilemma has not been matched by a comparable examination of NASA and its many works, some of which have astronomical connotations.

It would of course be folly to abandon the Hubble Space Telescope now that it is working superbly. But it would be an equal folly to close Kitt Peak simply because it is 20 years old. Why not take the knife to the space station instead? If the space station could be postponed for a single month, the saving in interest charges would provide a decade's worth of the modest increase over which the McCray panel has been drooling. A cost-conscious Congress should think on that. □

Names for hi-jacking

"Taxol" is a trademark now, but Bristol-Myers Squibb should return it to the research community.

EVERYBODY knows what is meant by the English noun "rock". It has three principal uses, in references to stones (that may be thrown), geological deposits (that may be laid down) and a form of rhythmic and cacophonous music, much enjoyed by young people. So what would happen if some recording company were able to proclaim that others should not use the term in any of those meanings on the grounds that "Rock" had become the trademark of its own brand of music? There would, of course, be a revolution. What geologists would do, and to whom, is too dreadful to imagine.

Sadly, there will be no such reaction to the exercise of similar claims by Bristol-Myers Squibb, the US pharmaceutical manufacturer, which has appropriated the word "taxol" for use as a trademark, with the willing but naive consent of the OPT. When *Nature* last year published an account of the total synthesis of this antileukaemia substance (Nicolaou, K. C. *et al. Nature* **367**, 630–634; 1994), a vice-president of the company wrote to say that "taxol" is a proprietary name and should not be used generically and, that "paclitaxel" is the approved alternative.

For the past two decades, the word "taxol" has been a generic name for a material extracted from the bark of the Pacific yew. It was so used in a paper describing the isolation and structure of the compound in 1971 (Wani, M. C. *et al. J. Amer. Chem. Soc.* **93**, 2325–2327; 1971). Bristol-Myers appears to have inherited its trademark from Continental Laboratories Ltd, which registered the name in 1939 to describe a "pluriglandular product to regulate the intestines". In 1992, Bristol-Myers applied for (and was granted) protection of the same name to refer to its "anti-cancer preparations", making a passing reference to the earlier protection for something quite different.

Bristol-Myers should be ashamed of itself. If it values its relationships with the research community, it should voluntarily relinquish the protection it has won "in more than 60 countries". Meanwhile, the US president or one of his underlings might usefully enquire why the trademarks examiners were asleep in 1992. □

Smithsonian to study museums' role after dropping A-bomb exhibition

Washington. Michael Heyman, the secretary of the Smithsonian Institution in Washington, announced on Monday that the institution is to mount a joint forum with the University of Michigan to look at how museums handle controversial topics — and in particular at the role of the Smithsonian as a national museum.

The announcement coincided with a decision by the Board of Regents of the Smithsonian, after a year-long battle between war veterans and the museum, to cancel an exhibition devised as a vehicle to display the *Enola Gay* — the plane that dropped an atomic bomb on Hiroshima on 6 August 1945.

In place of the planned exhibit, which had provoked fierce protest over conflicting interpretations of the historical significance of the event, only the plane and a video about its crew will now be shown. The Smithsonian says it is considering a series of symposia tackling the historical issues surrounding the *Enola Gay* at a later date.

Presenting the board's decision, which was reached unanimously, Heyman said he felt that it had made a "basic error" in trying to link together an historical treatment of the use of atomic weapons with an exhibition designed to commemorate the end of the Second World War.

The board, of which Heyman is secretary, includes members of both the public and of Congress. It had come under pressure from many sides. Last month, for example, 81 members from both sides of the House had called for the resignation of Martin Harwit, director of the National Air and Space Museum, which is responsible for displaying the *Enola Gay*. The board did not back the call.

Previously, veterans' groups had called through organizations such as the American Legion for the cancellation of the exhibit, despite expressing a desire to see the plane on display. But the Organization of American Historians wrote to William Renquist, the board's chairman and US Chief Justice, claiming that cancelling the exhibit would set a dangerous precedent of censorship under pressure from special interest groups.

Speaking after the meeting, Senator Dan-

iel Patrick Moynihan (Democrat, New York), a board member, emphasized that it had been unanimously decided to replace the exhibition with something less judgmental.

Although the board has now made its views known on the proposed *Enola Gay* exhibit, the issues raised by the controversy over its portrayal of one of the defining events of the twentieth century is set to continue. Such issues include questions about the use of history in American politics and the extent to which ideology can colour a



museum exhibit.

Other issues range from the difference between memory and history to the role of a national museum versus one that is privately funded, and the role which the curator of a museum plays when displaying artifacts — as opposed to that of a scholar publishing work under his or her own name.

The controversy, which has included charges that personal ideology plays an excessive role in some curatorial decisions, has already led to the postponement of an exhibit about Vietnam. It has also contributed to Heyman's decision to appoint an independent panel to evaluate another controversial exhibit *Science in American Life*.

The latter exhibit, mounted by the National Museum of American History — another of the Smithsonian's complex of museums — has come under fire from many

scientists who have complained that it emphasizes the negative consequences of science on society, and underestimates the positive contribution.

In an article due to appear in *Science Communication*, published by the American Association for the Advancement of Science (AAAS), Marcel LaFollette, an associate professor at George Washington University, expresses concern at the way retrospective moralizing appears to be creeping into some exhibits.

The *Enola Gay* has been in the Smithsonian collection since 1949. Restoration began 11 years ago, and will have eventually cost the museum \$1 million. The complete aircraft will go on permanent display at the National Air and Space Museum's new site, currently being developed at Dulles International Airport outside the national capital.



The *Enola Gay* (top) returns from its mission. Its pilot Paul Tibbets (above) has joined critics of the proposed display.

Meanwhile the museum has already moved the forward fuselage of the plane into the museum where it sits shrouded, awaiting an unveiling in May. (The weight of the complete plane exceeds the museum's load-bearing design specifications).

Prior to this week's cancellation, the exhibit had passed through a number of incarnations, originating several years ago with the idea of examining the events surrounding the dropping of the atomic bomb in the context of the start of the Cold War.

But some observers say that Harwit was equally keen to see an exhibit placing the plane in the context of the end of the Second World War. The decision to create two foci for the exhibit seems to have caused much of the subsequent trouble — but not all.

Early scripts contained a sentence saying that for most Americans it was a "war of vengeance", whereas "for most Japanese, it was a war to defend their unique culture against Western Imperialism." Whatever was intended by those words, veterans' groups have interpreted them to mean that they are racist avengers.

Much of the subsequent rhetoric has focused on the casualty estimates for an invasion of the mainland that military advisers gave to President Harry Truman.

Helen Gavaghan

NIH selects new alternative medicine head

Washington. Wayne Jonas, the director of the Medical Research Fellowship at the Walter Reed Army Institute of Research in Washington D. C., is to be the new head of the Office of Alternative Medicine at the US National Institutes of Health. Jonas, who takes up his post in July, has worked in immunology and toxicology, and has a

background in homeopathy.

Marc Micozzi, director of the National Museum of Health and Medicine, says that Jonas will be "excellent" in the post. Micozzi is also editor in chief of the *Journal of Alternative and Complementary Medicine*, a peer-reviewed journal which publishes its first issue this week. □

US accused of 'cover-up' in defence of Gallo claims

Paris. The US administration engaged in a massive 'cover-up' during the 1980s to defend the claims of Robert Gallo of the National Institutes of Health (NIH) that he discovered the AIDS virus used to develop the US HIV test, according to the draft report of a three-year investigation into the affair by staff members of the US Congress.

The report was prepared for John Dingell (Democrat, Michigan), who chaired the House of Representatives subcommittee that oversees NIH until the Republican majority took control of Congress last month. This changeover means that the draft report is unlikely to be released as an official subcommittee report. The Republicans have new concerns, and the Gallo affair is said not to be one of them.

The report contains a detailed study of allegations that Gallo committed scientific misconduct, and an investigation of the 'institutional response' to the HIV blood-test patent dispute. The United States and France brought about a ceasefire in this dispute in 1987 by agreeing to a settlement under which Gallo and Luc Montagnier of the Institut Pasteur were credited as being the 'co-discoverers' of the virus, and royalties from the sales of the test were shared among the US Department of Health and Social Services (HHS), the Institut Pasteur and the World AIDS Foundation (see *Nature* 326; 584; 1987).

But the report says that this settlement "barely managed to paper over the glaring unresolved issues". Rather, it was the culmination of a cover-up where "political and international reputational [sic] imperatives" at HHS "assumed pre-eminence over scientific integrity", while defending Gallo's claims became "tantamount to defending the US government itself".

Last summer, the United States agreed to

give the Institut Pasteur a greater share of the royalties, and Harold Varmus, the director of NIH, acknowledged that "scientists at NIH used a virus provided to them by Institut Pasteur to invent the American test kit" (see *Nature* 369, 693; 1994).

In contrast, the report claims that the HHS's response during the mid-1980s to allegations by scientists at the Institut Pasteur that Gallo had used their virus to develop the US AIDS test was intended "to defend, at all costs and irrespective of the evidence, the claims of Gallo *et al.*"

HHS carried out a "parody" of an investigation, the report alleges, seeking not the truth but to "create an official record" to support Gallo. In particular, the report criticizes an HHS investigation carried out by Peter Fischinger, one of Gallo's superiors, which, it says, relied exclusively on evidence from Gallo. "Contrary facts or evidence were neither sought nor examined."

The report also alleges that Bernardine Healy, the former director of NIH, "did everything she could to protect her superstar senior scientist" and "effectively demolished" the agency's Office of Scientific Integrity (OSI) (Suzanne Hadley, who left the OSI in 1991, became the principal staff member responsible for the investigation by the Dingell subcommittee, and is the main author of the draft report.)

The appointment of Varmus in 1993 ended "the atmosphere of overt protection" of Gallo, says the report. But it also criticizes Varmus for failing to pass on evidence "demonstrating probable misconduct" to the appropriate disciplinary body, the Surgeon General's Board of Enquiry.

Gallo's attorney Joseph Onk dismissed the report on Monday as an "unofficial document", and described its contents as meaningless.

Declan Butler

UK firms buy into drug design skills of US start-ups

London. Two British companies have taken advantage of the relatively low price of US biotechnology shares to acquire west-coast companies that will help their plans to use advanced computing techniques to design new drugs.

Last week, the pharmaceutical company Glaxo plc, fresh from its \$9.4-billion (US\$15-billion) bid for its rival Wellcome plc which would create the world's largest drugs company, announced that it had agreed to purchase Affymax, a company specializing in so-called 'combinatorial chemistry', for the sum of \$533 million.

A week earlier, in a smaller but also potentially significant deal, Oxford Molecular, a company specializing in computer-aided molecular design, which last year purchased the US company IntelliGenetics, announced that it had agreed to buy CAChe Scientific Inc. of Beaverton, Oregon.

Both moves indicate keen interest on the part of the British companies in becoming key players in the rapidly expanding use by pharmaceutical researchers of computer-designed molecules, rather than those produced through chemical experimentation, as the basis of a new generation of drugs.

Among the drug-discovery technologies developed by Affymax is one that uses silicon wafer coated with human genetic material — known as DNA chips — to analyse nucleotide sequences. This technique, which can potentially be used to screen for possible mutations of disease genes, has been commercialized by Affymetrix, a subsidiary company in which Affymax has a 65 per cent share.

Affymetrix scientists last year described as one of their goals the representation of a 10 kilobase gene as an array of 10,000 overlapping oligonucleotides on a chip, and thus as a means of the identification of mutants in a single pass (*Nature* 364, 556; 1994). The authors were then working with 10^5 synthesis sites per cm^2 , but were expecting a 10-fold improvement to be feasible.

Oxford Molecular's planned acquisition of CAChe will be discussed at a meeting of the company's shareholders next week. The company is based at the science park of the University of Oxford, which is also one of its main shareholders. It operates primarily through contracting out research to university-based researchers.

According to the company's chief executive officer, Tony Marchington, the purchase of CAChe, whose present owners are Sony of Japan and Tektronix of Oregon (both of whom will become major stakeholders in Oxford Molecular), will make the Oxford company Japan's largest supplier of molecular modelling software

Klug to take over at Royal Society

London. **Sir Aaron Klug (right), director of the UK Medical Research Council's Laboratory of Molecular Biology (LMB), has been nominated as the next president of the Royal Society, to succeed his fellow Cambridge academic, Sir Michael Atiyah, next November.**

Klug is widely known for his work on crystallographic electron microscopy and the structure of biologically significant molecules, for which he was awarded the Nobel prize for chem-

IMAGE
UNAVAILABLE
FOR
COPYRIGHT
REASONS

istry in 1982. Like his predecessor, he will hold the Royal Society post for five years.

Klug has also been a strong proponent of the need to protect basic science from excessive government influence. Two years ago, his stout defence of MRC laboratories such as the LMB helped to persuade the government to distance itself from proposals that all

such laboratories should be separated from the councils that finance them (see *Nature* 360, 614; 1994).

Kobe disaster divides earthquake researchers

Tokyo. A vigorous debate about the future of Japan's earthquake research — and in particular the earthquake prediction programme, in which Japan has invested substantial amounts of money and manpower over the past 30 years — has been generated by last month's Kobe disaster.

The Prime Minister, Tomiichi Murayama, opening the Diet (parliament) on 20 January, hinted that the prediction programme will be made "more comprehensive" and "nationwide". But earthquake researchers outside the programme say that the experience of the Kobe earthquake indicates that more money and effort should be put into other areas, such as the observation of strong ground motion and basic seismological research.

Citizens and government officials in Kobe say they had no idea that an earthquake of the magnitude of that on 17 January could hit their city. Some researchers claim that this was because too much attention has been focused on the Kanto and Tokai regions near Tokyo, lulling other regions of Japan into a sense of false security.

The Japanese government has designated 10 regions for extensive monitoring of earthquakes. One covers the cities of Kobe, Osaka, Kyoto and Nagoya, and the tip of Awaji Island, where last month's disastrous earthquake was centred.

But the southern Kanto region around Tokyo, as well as the adjoining Tokai region encompassing the Izu Peninsula south of the city, have been subject to even more intensified observations, as this is widely believed to be an area where major earthquakes are likely to occur in the near future.

Kiyoo Mogi, chairman of the Coordination Committee for Earthquake Prediction, and other leaders of the earthquake prediction programme, have repeatedly claimed in public that it may be possible to predict the next major earthquake in the Tokai region. Indeed, the government passed the Large-scale Earthquake Countermeasure Act in 1978 on the premise that the Tokai earthquake can be predicted and that, if precursors are detected, the Prime Minister will issue a public warning on the recommendation of Mogi's committee.

Such moves have focused the attention of both the media and the central and local governments on the Kanto-Tokai region over the past two decades. But Harumi Aoki, head of the Coordination Committee for Earthquake Prediction Research in Universities, denies that the result has been to mislead the public.

"We have repeatedly warned that the area [around Kobe] is riddled with active faults," he says. "But in practice only a few earthquakes have been felt, and the danger was therefore not appreciated." He and Mogi point out that a great deal of research has

been carried out in the Kobe area. For example, Japan's only 'test field' for earthquake prediction is on the Yamazaki fault, adjacent to that which ruptured during last month's earthquake.

Japan's emphasis on such prediction research is "lopsided" say earthquake engineers who would like to see more research on strong ground motion, where data can provide an instant measure of the scale of a

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Data on strong ground motions might help explain this structural failure.

disaster in different locations; they also provide important information on the movement of faults, and are needed to improve earthquake-proof designs for buildings and other man-made structures.

Japan has more than 2,000 seismographs for observing strong ground motion operated by national research organizations, universities and private companies. But according to Kenzo Toki, a civil engineer at Kyoto University, most are more than 20 years old and use wax paper and a stylus. Only a few transmit data electronically and most lack a common time axis. This contrasts with the networks for earthquake prediction and tsunami warnings, which are monitored continuously by the headquarters of the Japan Meteorological Agency.

Data from the strong-motion seismographs are collected by the National Research Institute for Earth Science and Disaster Prevention (NIED) in Tsukuba science city northeast of Tokyo. But NIED receives only about ¥20 million a year for this research, compared with about ¥4,000 million a year for earthquake prediction research.

Rapid release of data is essential for assessment of damage immediately after earthquakes. But NIED will not be releasing its first "prompt" report on strong ground motion of the Kobe earthquake for about a month.

Aoki opposes diverting money for prediction research to observations of strong ground motion, pointing out that the budget for disaster prevention is much larger than that for earthquake prediction. "Why not

discuss how to use the disaster prevention budget?" he asks. His view is echoed by Mogi, who says that the budget for earthquake prediction research over the past decade "would not even pay for 1 km of road" around Tokyo.

But funds for university research on disaster prevention are limited. In 1994, only 0.1 per cent of the ¥3,342 billion spent by the government on natural disasters was allocated for research in universities. The Ministry of Education, Science and Culture has no category of research grants suitable for establishing strong-motion observation networks, says Toki, forcing him to turn to the private sector for help when setting up a telemetered network of 10 seismographs in the Kobe-Osaka area last year.

In the United States, seismologists and earthquake engineers frequently work closely together. For example, over the past 5 years they have collected both microquake and strong-motion data from seismograph networks such as the California Institute of Technology's TERRAscope and the Caltech-USGS Broadcast of Earthquakes (CUBE). In Japan, earthquake engineers and seismologists are still "on opposite banks of the river", says Toki, claiming that most seismologists have no interest in strong ground motions that cannot be used for prediction research.

Others complain that the emphasis on prediction research has led to the neglect of basic seismological research. Robert Geller, for example, a geophysicist at Tokyo University, has been campaigning for the replacement of the prediction programme with a new programme that would include more fundamental research aimed at understanding the structure of the Earth (see *Nature* 352, 275; 1991).

Aoki points out that the prediction programme has led to many fundamental advances in the understanding of the Earth, such as the discovery of the double seismic zone in the subduction zone off the Pacific coast of Japan. Geller says he accepts this example but points out that it was not the aim of the research.

But Aoki and Mogi are both adamant that research on earthquake prediction should be expanded. Given the new concern for inland earthquakes, "observations will be strengthened, never weakened" says Aoki. Mogi adds that "even if other countries give up earthquake prediction, Japan should continue; there is no other country that can deal with this thoroughly".

What changes the government will make, if any, remains unclear. At present, its main advisers are those involved in earthquake prediction research. Unless it consults more widely, a radical redirection of Japan's earthquake research seems unlikely.

David Swinbanks

US technology support under the microscope

Washington. The Advanced Technology Program (ATP) — the programme closest to the heart of the Clinton administration's science and technology policy — went on trial this week, with officials from the Department of Commerce summoned to Capitol Hill for the first of a series of hearings at which opponents will press for its elimination.

Industry leaders have urged Congress to ensure that the trial is fair and to assess the programme on its merits. But recent Republican rhetoric, as well as earlier, highly partisan debates in the House of Representatives, have branded it as 'industrial policy', to which the new majority in Congress is strongly opposed.

Ron Brown, the commerce secretary, Mary Good, his under-secretary for technology, and Arati Prabhakar, director of the National Institute of Standards and Technology (NIST), which is responsible for running the ATP, were expected to offer an energetic defence of the programme at a hearing of the science, technology and space subcommittee of the Senate Commerce Committee, which was due to be chaired by Senator Conrad Burns (Montana) on Tuesday (31 January).

The Clinton administration has expanded funding for the programme, which helps industrial companies with pre-competitive research and development projects by paying half their costs, from \$68 million in 1993 to \$199 million in 1994 and \$431 million this year.

Such increases have made ATP by far the fastest-growing element in a federal science and technology budget that has remained flat overall at around \$73 billion a year — and an obvious target for Republi-

can cuts. "It's probably the most vulnerable programme in science and technology," says David Goldston of the Council on Competitiveness, a centrist Washington think-tank.

Republican opponents, led by Robert Walker (Republican, Pennsylvania), chairman of the House Science Committee,

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Light work: ATP projects include use of hologram technology for fibre optics.

have singled out the ATP for attack since last November's elections. Staff members say that the main issue is whether the programme should be severely curtailed, or closed altogether.

But Prabhakar said last week that her own conversations with congressmen were "going very well" and that the programme still retained a measure of bipartisan support. "We've heard a lot of rhetoric from people that don't know anything about how the programme works," she says. "My concern is to make sure the facts are on the table."

The programme, Prabhakar points out, began under the Bush administration after years of effort by industrial groups and government officials to decide how govern-

ment could best help US industry with technological innovation. "ATP was designed on the basis of literally decades of false starts, with lots of smart people thinking about how to make it a really effective investment," says Prabhakar.

Its fate, she concedes, now rests with the willingness and ability of those whom ATP is supposed to benefit to lobby on its behalf. "The real issue will come down to whether industry thinks these are the right programmes."

Although the 1988 authorization of the programme and its first money — \$10 million in 1990 — were both driven by what was then a Democrat-dominated Congress, the programme's strongest early sponsor was Don Ritter, a since-defeated Republican Congressman from Pennsylvania. D. Allan Bromley, Bush's science adviser, became an enthusiastic proponent of ATP.

Bromley, now dean of engineering at Yale University, Connecticut, says that the ATP deserved to be expanded rapidly on the basis of its early performance. "Its first year was a great success," he says. "The fundamental problem is that the Clinton administration chose to make it a showpiece of their industrial policy." Bromley takes a more sanguine view than most of the prospect for ATP's survival. "When the dust settles, funding will be cut back somewhat but the programme will remain".

Even some of ATP's supporters feel that it was allowed to grow too quickly. The administration made a mistake in expanding ATP so fast, says William Morin, director of technology policy at the National Association of Manufacturers (NAM), a powerful lobby group. "It hadn't been around long enough to establish a track record."

NIST has plenty of case studies to talk about, but no real economic data on its success. Prabhakar says it will take 10 or 15 years for such data to emerge, and NIST has therefore had to use "short-term impacts", such as new business alliances and participants' impressions, to assess the programme. The NAM and other industry groups still support the Advanced Technology Program, and Morin says Congress should assess its effects "rather than subjecting it to some sort of ideological litmus test".

Last year, Clinton proposed the further expansion of the ATP to \$540 million in 1996 and \$770 million in 1997. His 1996 budget proposal, due to be published next week, may aim lower: the real budget, set by Congress over the summer, much lower still. Morin worries that there will be no real discussion of technology policy in the process. "We ought to go about this in a much more measured way," he says.

Colin Macilwain

'\$1 a shot' goal for DN374A analysis

Washington. As well as single-company projects worth up to \$2 million, the Advanced Technology Program (see above) supports consortia engaged in complex interdisciplinary innovations. Eight partners, for example, are engaged in the Genosensor Consortium, which, according to Stan Abramowitz, NIST programme manager, aims to cut the costs of DNA analysis by a silicon chip from more than \$100 a test to \$1.

The Genosensor project, whose costs of \$18 million over five years will be shared equally between NIST and industry, was conceived by Beckman Instruments, a laboratory-equipment company based in Fullerton, California. It will involve chemists and molecular biologists at Baylor College of Medicine, in Houston, Texas, and elec-

tronic engineers at the Massachusetts Institute of Technology, as well as researchers at the Houston Advanced Research Center and five industrial companies, to establish how such a machine could be built.

"We're trying to automate the techniques of molecular biology and make the devices small, simple and easy to use," says Jim Osborne, vice-president for advanced chemistry at Beckman. No one partner could have carried the risks inherent in the project, he says. Like other ATP projects, the year-old Genosensor could be killed if the programme is cut back sharply this year: like all government grants, NIST's promised contribution is contingent on adequate funds being available. **C.M.**

Fusion project 'should wait for US results'

Washington. The International Thermonuclear Experimental Reactor (ITER) should be delayed for between three and five years to allow the United States first to fund its own fusion experiment, the Tokamak Physics Experiment (TPX) at Princeton, New Jersey, according to a leading congressional supporter of the fusion programme.

The comments come shortly before the release of a report from the congressional Office of Technology Assessment (OTA), which is expected to castigate the US Department of Energy (DoE) for the lack of coordination between the two main magnetic fusion programmes, ITER and TPX, and to suggest that TPX would be a far more useful experiment if ITER were delayed to incorporate its results.

ITER, which is being planned jointly with research teams in Europe, Japan and Russia, is now in the design stage, with construction scheduled in principle to begin in 1998. But George Brown (Democrat, California), the ranking minority member of the House of Representatives Science Committee, and one of ITER's strongest backers in Congress, said last week that the possibility of a delay should be examined before going ahead with construction. "We should examine that with our international partners. I can foresee a pause of three to five years."

OTA's unpublished report, *The Fusion Energy Program: The Role of TPX and Alternative Concepts*, says that the value of TPX to the magnetic fusion programme "could increase" if ITER were delayed. At present, it says, the planned schedule of the two projects prevents TPX results from feeding into the design and construction of ITER. This could only happen, it says, if ITER construction were delayed for "several years".

Robert Aymar, director of ITER, immediately rejected the idea of a delay to wait for TPX. "It's completely out of the question," he said. According to Aymar, TPX would not yield results until 2005, and a delay until then "was like saying not to continue ITER at all".

TPX is designed to test a number of innovations in fusion reactor design, including the use of superconducting magnets and, most importantly, the use of a controlled current profile to sustain fusion in a so-called "second stability regime" of higher density than has been managed before.

Stephen Dean, president of Fusion Power Associates and an acknowledged fusion expert, claims that the construction of an ITER prototype would be smaller, and 20–40 per cent cheaper, if built on the basis of a successful TPX experiment.

The OTA report also disputes the DoE's contention that it needs \$100 million to carry out research into alternatives to tokamaks, the doughnut-shaped fusion chambers on which TPX and ITER would be based. It says that useful paper studies of

alternatives could be conducted for only a few million dollars.

The postponement of the ITER would relieve pressure on the department's magnetic fusion research budget — \$373 million this year — which is too small to support ITER, TPX and the supporting basic research programmes (see *Nature* 372, 493; 1993).

But the proposal will raise difficult new questions, such as whether the money now being spent on ITER design, planned at \$150 million this year in the United States alone, is being wasted if the project is delayed and then redesigned.

President Bill Clinton's budget proposal for the fiscal year beginning on 1 October,

due to be revealed next week, is likely to continue funding for fusion at roughly the present level. But administration officials have said that their plans will be revised in the light of two reports due in the spring — one on research priorities from a DoE advisory panel, the other on fusion from the President's Committee of Advisors on Science and Technology (PCAST).

The attitude towards fusion of the new Congress, where budgets will be finally set, remains untested. But one certainty, as Dean points out, is that dollars will be even harder to find for an international project such as ITER than for the all-American TPX.

Colin Macilwain

Threat to German medical research

Munich. Medical research is suffering because of the priority being given to short-term patient care, according to Germany's science council, the Wissenschaftsrat. In a report published last week, it says that major changes must be made in the management of teaching hospitals to increase the decision-making powers of academic faculty and ensure that funds intended for research are not diverted to patient care.

Most teaching hospitals operate with a single budget for patient care, teaching and research, and distribution of funds between these functions is often obscure. The Wissenschaftsrat wants accounting to be more transparent, and teaching hospitals to maintain separate budgets for teaching and research.

In many cases, decisions about the distribution of funds are made by the clinical faculty without consulting academic staff, says Gisela Frank, head of the Wissenschaftsrat's medical research team. "When money becomes tight, this creates a conflict between patient care and research and teaching," she says.

The problem is particularly pressing because of fears that a new health-care law will lead teaching hospitals to cut funds for research. Under the new law, which has already frozen funding at 1993 levels and comes into full force in 1996, health insurance will no longer pay the hospitals' actual costs, but only the estimated average cost of each patient procedure. This, it is hoped, will encourage hospitals to keep costs down.

But many fear that the move will penalize teaching hospitals, which receive between 60 and 90 per cent of their funding from national insurance, and the rest from local governments. As they treat unusual and severe cases, and require more staff to train new doctors, teaching hospitals are often more expensive than

community hospitals. If they are to receive less money from health insurance, researchers fear that clinicians will fight for a higher share of the local government money that supports research.

But separating expenditures is not easy, says Klaus Peter, dean of the medical school at the Ludwig Maximilian University in Munich. "Clinical costs represent 30 to 40 per cent of the operating budget, and research and teaching 20 per cent, but it is impossible to separate and specify the rest."

For example, while facilities such as outpatient clinics are part of patient care, they are also essential for training doctors. Peter Gaehtgens, a physiologist and, until recently, vice-president of the medical faculty of the Free University of Berlin, says that national health insurance does not recognize this overlap, nor does it cover such services adequately. The result is that less money remains for research.

The Wissenschaftsrat also recommends closing or combining some facilities. In particular, it suggests that the two medical schools in Munich should combine some of their facilities, and that 600 beds from Munich and other hospitals in Bavaria should be closed to free funds for research. "This will not harm patient care, since more procedures are being handled on an outpatient basis", says Gaehtgens.

The recommendations of the Wissenschaftsrat, whose members include both scientists and politicians, are influential at the federal level. Although teaching hospitals are almost entirely locally funded, half of the costs of new buildings and major items of equipment are paid by the federal government — which is why the hospitals are already starting to implement the advisory group's proposals.

Toni Feder

Patent system gets vote of support from gene workers

Paris. The patent system is the 'mechanism of excellence' for commercializing the results of the human genome project. This was the unexpected consensus reached at an international meeting of leading genome scientists, industrialists and patent lawyers organized by the French Academy of Sciences in Paris last week.

The mood of the meeting was that the concern about patenting genes had now been allayed. "Patenting genes poses no specific problems," said Philippe Lazar, director general of the French biomedical research organization INSERM. But it had helped to prompt an overdue debate on the role of patents in research, technology transfer and the sharing of *savoir-faire* with developing countries.

"Patents are the instruments of partnership between science and the economy," said Lazar, summing up an enthusiasm for the patent system among researchers at a meeting which François Gros, the perpetual secretary of the academy, pointed out would have been unthinkable 15 years ago.

In particular, many felt that most of the current problems associated with patenting genes would disappear over time, particularly if patent offices simply applied the three existing criteria of patentability — novelty, non-obviousness and utility — with greater rigour. Indeed, the meeting concluded that commercializing the genome does not require reinventing the internationally proven, 200-year-old patent system, but simply adapting it.

One major concern, said Thomas Caskey, president of the Human Genome Organization (HUGO), is that patenting partial or uncharacterized cDNAs "would reward those who make routine discoveries, but penalize those who determine biological function or application". This is not in the "public interest" he said.

"We've got to explain to lawyers that finding a gene [using routine techniques] is not inventive," said Jean Weissenbach from the French genome centre Génethon. "We all agree that genes can be patented when we know their function," added Maxime Schwartz, director general of the Pasteur Institute.

Patent lawyers told the meeting that patents whose function is unknown are unlikely to be refused on the grounds of lack of novelty. This criterion can be easily sidestepped, for example, by describing genes as "isolated, purified, and carried in a recombinant host". But they said that patents on such genes could be refused for lack of inventiveness — or non-obviousness — in particular as more and more genes are isolated.

Another related concern expressed at the meeting was the fairness of awarding a patent to the research group that places the

final brick, for example the actual discovery of a specific gene, in a wall built by many others. Patent lawyers pointed out however, that the patent system exists only to promote innovation and not to guarantee equity among inventors.

But Schwartz argued that "we have to reduce the disparity between those who get the patent and those who did the work". One possibility would be royalty-sharing agreements. Schwartz proposed, for example, that such agreements are particularly needed in areas such as AIDS vaccines, where the low potential return, combined with the all-or-nothing nature of competition, discourages investment.

But the lawyers dismissed researchers' concern that the granting of a patent to the discoverer of a gene would penalize those who subsequently carried out the more difficult task of working out what the gene does, and what it could be used for. They argued that there is no reason why a scientist should not obtain a separate patent on further uses of a gene, in much the same way as patent rights are now conferred on new applications of existing drug molecules.

Indeed, the meeting felt that recent attention given to the patentability of genes *per se* has detracted from the bigger problem of how such rights are enforced. The breadth of patents granted on gene technologies, and the terms under which patents are licensed, were identified as key issues.

Peter Kolken, for example, from the British pharmaceutical group Zeneca, argued that it was "normal" for patent offices to grant broad patents in the early days of a technology, and then to reduce the scope as the technology developed. Another patent lawyer argued that the only way to find the real scope of a patent is to challenge it in court.

Moreover, even if patent offices grant excessively broad patents, the patent system can handle this, according to other patent lawyers. There is no reason why compulsory licences, for example, should not be invoked more often for important inventions concerning the genome.

Caskey also asked whether patents on genes might increase health-care costs unnecessarily and restrict access to therapies developed from the human genome project. He suggested that governments should intervene to regulate the level of royalties.

Jean Lunel, a research director at Rhône-Poulenc Rorer, asserted that such arguments show a misunderstanding of the patent system. Although patents reduce access to therapies for part of the population in the short term, he argued, without patents the whole population would be deprived of new drugs.

Declan Butler

Terms of HGS deal boost search for new strategies

Paris. If patents are becoming more acceptable as a way of handling information about new genetic discoveries (left), this is partly due to reaction to the terms set by the US gene-sequencing company Human Genome Sciences Inc. (HGS) for university-based scientists seeking to work with sequences produced by HGS's Institute of Genomic Research (TIGR).

The agreement gives HGS first refusal on any patents arising from research using the database sequences (see *Nature* **371**, 463; 1994).

This requirement has generated widespread criticism. "We've discovered that there is something worse than patents", said Bertrand Jordan, from the Centre d'Immunologie de Marseille-Luminy in France, speaking at a meeting on intellectual property and the human genome in Paris last week.

Few researchers have accepted HGS's terms since they were announced in October. James Watson, the Nobel prizewinner and director of Cold Spring Harbor Laboratory, said, for example, that he could not enforce a confidentiality agreement demanded by HGS, that would restrict discussions between different research groups working on HGS data, and that this would expose his institute to legal action.

David Owen, from the UK Medical Research Council (MRC), says that the council considers that HGS's terms "restrain interaction" and has not authorized its researchers to sign contracts with HGS. Like many other laboratories, the MRC is trying to negotiate better terms.

But patents may not be the end of the story. This was illustrated at the meeting by Rebecca Eisenberg, from the University of Michigan Law School in Ann Arbor, who pointed out that "substantial private funds" are now being committed to cDNA sequences on which no patent rights are held. This, says Eisenberg, is providing a "real world laboratory" for viewing the competing strategies of HGS, Incyte Pharmaceuticals, and Merck Sharpe & Dohme for commercializing "unpatented" results from the genome.

D. B.



Watson: unwilling to endorse secrecy

Sparks continue to fly in *Taq* patent dispute

London. Cetus Corporation, the California-based biotechnology company, has been accused by the laboratory supply company Promega of presenting the US Patent Office with arguments that were "at best fallacious and at worst deliberately misleading" in defending its claim to a patent on the thermostable enzyme *Taq* polymerase.

The accusation is based on evidence that Promega is presenting to a federal court in California to back up a challenge to the validity of the patent, which was granted in 1990 (see *Nature* 372, 714; 1994). The case, including a response from Hoffman-La Roche, which purchased full rights to the patent in 1991 as part of a \$300-million agreement, will be heard in a San Francisco court next week.

It has also emerged that the Cetus/Roche claims on *Taq* polymerase are being challenged in Australia by another laboratory supply company, New England Biolabs, which sells its own thermostable polymerases VENT and Deep VENT. The main thrust of the challenge is over the breadth of the Australian patent; but a company lawyer said last week that Promega's position "is consistent with the position that we have taken".

Roche is planning to react vigorously in court to the charges which, in an unusual move, Promega has decided to publicize through making supporting documentation available directly on the Internet.

Tom White, vice-president for research and development of Roche Molecular Systems, said on Monday that the company, which normally has a policy of not commenting publicly on on-going litigation, intended to refute each of Promega's charges. "We are fully confident that our *Taq* patents will be upheld," he said.

The stakes on both sides are high. If Promega loses its challenge, it would seriously weaken its defence against a previous suit filed against it by Roche — also to be heard in court next week — for allegedly infringing a licensing agreement by permitting its (Promega's) *Taq* polymerase to be used for polymerase chain reaction (PCR) procedures.

The principal issue at stake remains whether the thermostable enzyme extracted from *Taq* polymerase by scientists working for Cetus is identical to similar enzymes extracted from the same organism — originally discovered in 1969 — which had already been described several times in the literature. If the enzymes were identical, then the existence of so-called 'prior art' would invalidate the Cetus patent.

Three specific instances are referred to by Promega. The first concerns work by a group of Russian scientists (Kaledin *et al.*) published in Moscow in April 1980, and in English the following year. The second is similar work carried out by a group at the

University of Cincinnati (Chien *et al.*) and published in 1976. Most recently, Promega has produced a PhD thesis published two years earlier by David Edgar, later a member of the Cincinnati group.

All describe the extraction of an enzyme with thermostable properties. The first two in particular were raised as possible objections by the US patent official who examined the original patent claim. In reply, Cetus claimed that the enzyme its scientists had 'discovered' was substantially different, both in terms of molecular weight and in the accuracy with which it reproduced nucleotide sequences.

The patent was granted on the basis of these replies. But Promega is now claiming that a series of experiments it has either carried out or commissioned on the Edgar, Chien and Kaledin enzymes indicate that all three are identical to the Cetus enzyme — throwing doubt on the validity of Cetus's claims to the contrary.

One test, for example, carried out by Randall Diamond, chief technical officer of the company, using a Western blot technique with antibodies specific for *Taq* polymerase, showed virtually identical results for the molecular weight of approximately 86,000 to 90,000 daltons of the Edgar and Chien enzymes when compared to the Cetus enzyme (see figure).

Promega claims that other tests challenge claims by Cetus that the polymerase preparations described in the earlier references had been degraded during their purification, and the fidelity of their own enzyme was superior to that of preparations described in these references.

Similarly, Cetus/Roche have placed considerable weight on the results of a preliminary analysis by Jerilyn A. Verhoeven, of the University of Cincinnati, which appeared to show that the Chien enzyme "is actually composed of two identical subunits each

having a molecular weight of 35 kd". It used this argument, for example, to respond to criticisms from both the US Patent office and the European Patent Office (EPO).

But Promega says that subsequent experiments designed to verify this work (whose results were published in a peer-reviewed journal) suggest that this conclusion is flawed, as they showed that the 35-kd protein described by Verhoeven in a "preliminary" report "is not a polymerase" — and that the polymerase she had in fact purified had a molecular weight of approximately 90,000 daltons.

Promega's conclusions have been endorsed by Arthur Kornberg, professor emeritus of biochemistry at Stanford University — and winner of the Nobel prize for medicine in 1959 for his work leading to the discovery of DNA polymerase.

In a declaration supporting Promega's challenge to the *Taq* patent, Kornberg says that he has reviewed all the evidence produced by the company to back up its claims. He concludes that the polymerase described by Edgar, Chien and Kaledin "is the same as" the polymerase described in the Cetus patent claim, and that "certain representations" made to the patent examiner concerning these earlier enzymes and their activity "were incorrect".

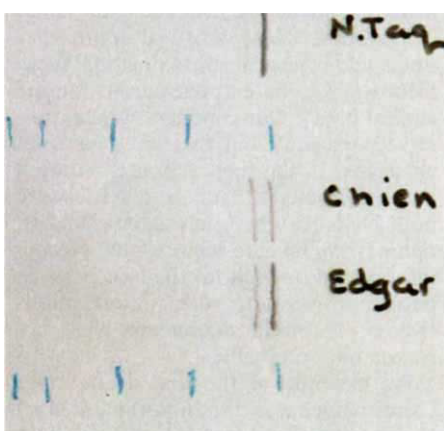
But Roche promises to contest these conclusions strongly in next week's court case. White said on Monday that the company intends to demonstrate to the court that "there are several key physical and biochemical characteristics that will distinguish our product from those in the prior art".

Roche also argues that the material which Promega has produced as evidence of "prior art" has already been considered — and rejected — by patent examiners in four continents, namely the United States, Europe, Japan and Australia.

Diamond says that his company's decision to issue details of its allegations on the Internet (they can be obtained through anonymous log-in on ftp.promega.com) is based on a desire to challenge Roche's assertions that its defence of the patent is secure. But he admits, "ultimately the issue is going to be decided in the courts".

Indeed, whatever the outcome of next week's hearing, the issue is likely to rumble on (keeping lawyers in business) for several years. On Monday, for example, a spokesman for the EPO, said that an EPO examiner, having previously told Roche that the patent was to be granted (see *Nature* 372, 212; 1994), has at the last moment decided to "reconsider" this decision in the light of new evidence presented by a "third party". Roche said it had not been informed of any such decision — and remained confident that the European patent would still be granted.

David Dickson



Valid proof? Promega's comparison by Western blotting of the *nTaq* enzyme with those in two earlier references.

Federations and republics

SIR — Switzerland is not a good example of federal dissolution (*Nature* 373, 89; 1995) and did not lose Jura to France. Rather, 69.9 per cent of the Swiss electorate having voted in favour, a new canton, Jura, was established *inside* the Swiss Confederation on 1 January 1979.

Your mention of Jura is nonetheless timely, and President Boris Yeltsin could if he wished still turn the Chechnia quagmire to his democratic advantage. All he needs to do is read a little history. The 'Jura question' goes back well before the period (late 1850s) when the Chechens were being conquered by Russia.

The Swiss Confederation, which has expanded steadily in the past 700 years, is very different from the Russian model. One grew by plebiscite and the putting of heads together, the other by invasion and the bashing of heads together. The former has lasted and matured on the basis of the powerfully attractive concept of self-determination, whereas dogma has turned the latter into "the carcass of a failed empire" (*The Economist* 334, 23; 1995).

There were four quite separate motives for the Chechnia intervention: politicians believed that secession could only sabotage their efforts to breathe new life into an old dream; hard-line strategists were determined to protect the oil pipeline between Baku and Novorossiisk; army commanders wished to give their conscripts something to do; and the Russian man-in-the-street felt that it was time to curtail the deadly activities of hired Chechen gangsters. But it all went wrong.

To the *nomenklatura*, who have been busy saying, 'You see, democracy doesn't work', Yeltsin can now retort, 'You see, tanks don't work; let us have a referendum'. And the tiny Republic of Chechnia, supported by the Russian electorate, could take its place proudly inside the new federation — as did little Jura, when it seceded peacefully from Berne, not so long ago.

A. L. T. McCammon
Tägernstrasse 20,
8127 Forch, Switzerland

SIR — When Flashman, in *Flashman and the Angel of the Lord* (by George MacDonald Fraser), asked Lincoln why it had been all right for the 13 states to secede from the British Empire in 1776, but wrong for the South to want to secede from the Union in 1861, Lincoln replied, "What had right to do with it? The first succeeded and the second failed." A fictional account that embodies an historical truth. History is largely about winners — as of course is science.

John Galloway
54 Windmill Drive,
Croxley Green, Hertfordshire WD3 3FE, UK

SIR — In a discussion of the implications, for science, of the Republican majority in the 104th Congress, Colin Macilwain writes (*Nature* 373, 5–6; 1995) about "the US Constitution, which was drawn up by Virginian landowners. . .". One expects better than such historical inaccuracy from *Nature*. The Constitutional Convention met in Philadelphia from 25 May 1787 (the first day when a quorum of seven states was assembled) until 17 September, at which time, of the 55 Framers, 41 were present, all but three of whom put their names to the Constitution. All 13 states except Rhode Island had sent delegates. Virginia sent seven, only some of them landowners.

The two committees most responsible for the actual writing of the Constitution were the committee of detail (one Virginian, Randolph, no landowner) and the committee of style (one Virginian, Madison, a man of such modest means that he had to be supported at the time by his father and friends). The consensus embodied in the final instrument was hammered out in discussions among the delegates during the hot summer of 1787.

So the statement that the Constitution "was drawn up by Virginian landowners" is a flagrant misstatement of the facts. The facts are readily available (see for example 1787 — *The Grand Convention*, Clinton Rossiter, 1966).

Emanuel Epstein
Department of Land, Air and Water
Resources,
Soils and Biogeochemistry Program,
University of California,
Davis, California 95616, USA

Randomized trials

SIR — In offering caution about a phrase about non-randomized trials in our Commentary¹, Peto and Collins² risk misleading readers. We said and continue to believe that "New statistical approaches often allow results approximating those from trials to be extracted from routine clinical data". Our comment was addressed to two situations that often confront clinicians. In the first, action is required and the results of randomized trials are not available. In such circumstances, optimal patient care requires that *existing* information be used for the best possible decisions. Second, when interventions may be harmful, randomized trials are clearly not acceptable.

An example of the first is the close correspondence of the (non-randomized) Duke coronary artery disease database and the results of subsequent randomized trials of coronary artery bypass grafting^{3–5}. Other methods for observational

studies can be improved to yield results that are more similar to those of randomized trials⁶.

We strongly agree, of course, that randomized trials are to be preferred when they are feasible and ethical. But alternative sources of information should not be dismissed out of hand. Physicians and policy-makers must often rely on non-randomized observations until the results of definitive trials are available. Finally, occasional interventions may have such a substantial impact that the moderate bias introduced by non-randomized methods can be tolerated. We should therefore continue to encourage the development of methods to use detailed routine clinical data, which should not be confused with the demographic information in an administrative database.

Lee Goldman
Howard Hiatt
Brigham and Women's Hospital,
75 Francis Street,
Boston,
Massachusetts 02115, USA

1. Goldman, L. & Hiatt, H. *Nature* 371, 100 (1994).
2. Peto, R. & Collins, R. *Nature* 372, 588 (1994).
3. Hiatt, M. A. et al. *J. Am. Coll. Cardiol.* 11, 237–245 (1988).
4. Yusuf, S. et al. *Lancet* 344, 563–570 (1994).
5. Mark, D. B. et al. *Circulation* 89, 2015–2025 (1994).
6. Horwitz, R. I. et al. *Am. J. Med.* 89, 630–638 (1990).

Popular science

SIR — One can only applaud and encourage the collaboration between *Nature* and *Le Monde* (*Nature* 373, 2; 1995). Popularization of scientific research is an valuable and important part of scientific work.

But I believe the flame of interest in the layman must be kindled at a far earlier stage than the average age of the readership of the broadsheets. As the product of the French schooling system and subsequent British university training, I fully agree with the comment that French people retain intellectual interest far beyond that of many nations. This I believe is as much due to their educational philosophy as to any other national trait, and it is symptomatic of that approach that a French newspaper has joined *Nature* in this venture. French education is based on instilling a sense of pride and importance in the pursuit of knowledge. An individual is believed to thrive on more than practical or physical skills. This interest subsequently remains throughout the individual's life.

Britain might benefit from considering this approach because it would, I believe, become far more receptive and sympathetic to the benefits and importance of science.

A. S. K. Dzilk-Jurasz
11 Carlton Park Avenue,
London SW20 8BJ, UK

Female ensoulment: late but durable

SIR — John Godfrey¹ writes with approval of the view expressed by Thomas Aquinas in the thirteenth century that “ensoulment”, the becoming of a human person, occurs gradually during embryonic and fetal development. The only point of disagreement is with Aquinas’s acceptance of the Aristotelian notion that ensoulment is completed at 40 days in male embryos compared with 90 days in female embryos, which seems an “exotic” error, based on misleading evidence about the development of males and females. I should like to point out, however, that the apparent bizarreness of this idea owes less to embryology than to the extreme reluctance of most present-day molecular biologists to take on board quantitative, as opposed to histological and biochemical, aspects of development, and to consider the importance of time.

A quarter of a century ago, Alfred Jost² stated that the first histological sign of male sexual development (seminiferous tubules) is detectable in human embryos at 6 weeks, and that of female development (primary oocytes) at 11 weeks. More recently it has been shown that male embryos develop faster than female ones right from the beginning³, and that male preimplantation cattle embryos have higher metabolic rates than their female counterparts⁴. Human males have higher metabolic rates than females throughout their lifetime⁵, and this could be a reason why males have a shorter life expectancy than females⁶.

Of course, Aristotle was no devotee of equal opportunities. He also had neither a microtome nor a microscope, and in any event would not have known the genetic sex of embryos. The ancient Greek view of the female as a somewhat underdeveloped male would have been based on observed facts, for instance that women have less muscular strength, no beard, and so on, which were interpreted with a good dose of chauvinism. Nevertheless, even if female ensoulment is delayed for 7 weeks, women would seem to be amply recompensed by the expectation that the reverse process of the soul’s departure will be delayed by at least 5 years.

Ursula Mittwoch

Department of Anatomy,
Queen Mary and Westfield College,
London E1 4NS, UK

SIR — Whatever reasons Aquinas had for dating the origin of the human person at 40–60 days after conception, no such reasons are available to those with the benefit of twentieth-century embryology.

The human person is (as Aquinas certainly recognized) an animal of a rational kind, and not a mere ‘soul’, or life-

principle. An animal is a living whole — to be distinguished from a living part, such as a gamete. There is no reason why the living whole should not be seen as originating on completion of sperm entry, when there is first present an entity with the active potential to develop as (not into) an individual human being.

Syngamy is no more than the rearrangement of chromosomes already present in the one-cell embryo. Gene expression is no more than the acting-out of tendencies already present in that embryo. As for the lack of individuality often inferred from the potential to absorb foreign genetic material, if this were a bar to individuality, blood transfusions — not to mention gene therapy — would retrospectively cast doubt on the individuality of the adult human being. As for the possibility of twinning, why should asexual, any more than sexual ‘reproduction’, be incompatible with the prior existence of the twinning individual?

Once we have identified the embryo as the same individual as the adult human being, we can recognize its interests in fulfilling its potential as the kind of animal it is. While we may have strong emotional grounds for denying that the embryo is a human person, the rational case in favour of this claim is distinctly unconvincing.

Helen Watt

Linacre Centre,
60 Grove End Road, London NW8 9NH, UK

SIR — While agreeing with Godfrey¹ that individual identity is acquired gradually and progressively, I am not sure this will help persuade the Pope to alter his advice to Roman Catholics. If one starts from the position that it is sinful to terminate life, there follows a need to define a point at which it begins. Three such points are in current use: fertilization, implantation and neurulation. The Pope (and others, Catholics and the rest) might reason in the face of the developmental continuum described by Godfrey that the safest strategy is to define the earliest possible stage — fertilization — as the point at which individuation is set in train. Implantation, when the maternal system begins active support to the conceptus and therefore recognizes its potential viability, is a conceptually well-defined alternative start point. In keeping with this, preimplantation ‘contraception’ is widely accepted as

ethically superior to abortion. The use of neurulation to define the onset of personhood may reflect the tendency of Western civilization to identify the brain with the individual. But this stage has the practical advantage that it occurs later, when pregnancy can be recognized by a urine test.

Biologists may differ as to the precise definition of these events, their duration or exactly when they are complete, but they remain conceptually well-defined developmental stages. As far as the lay person is concerned, debating whether fertilization takes a fraction of a second or two days is tantamount to counting the angels on a pinhead.

Godfrey’s assumption seems to be that the Catholic Church under Pope John Paul II can be engaged in a debate over reproductive issues. Before this can occur, the church will have to change its attitude to contraception rather than abortion. Given that many Catholics ignore the church on both issues, science should strive to provide them with the means to act in a manner that is increasingly ethically acceptable, as well as being compatible with the realities of unplanned pregnancy and the limited resources of the planet. This means earlier pregnancy diagnosis as well as a greater range of male and female contraceptive options.

John Aplin

Department of Obstetrics and Gynaecology,
University of Manchester,
St Mary’s Hospital,
Manchester M13 0JH, UK

Apoptosis forever

SIR — Søren Nørby (*Nature* 372, 312; 1994) wants the second ‘p’ in apoptosis silenced “to secure the understanding of the underlying chemistry”, with scant regard for the underlying physics of the screw-winged helicopter. Jonathan C. Busser and Alexandra Horowitz compare apoptosis and pterodactyl, to support their contention that “Funder misses an important distinction between pronunciation and spelling”. They justify apo’(p)ptosis on the grounds that “Because the term used to describe a particular type of cell death is derived from two greek words, apo- and -ptosis, the syllabic break is made between the two words when they are joined”.

This is clearly not the case. We say hypo’thesis, but hypo’thetical, breaking the word quite differently where balance and euphony dictate. For consonants the same considerations apply: we pronounce phthisis as tie’sis, but give the similarly greek derived ϕ and θ full play in ophthalmologist. *Archaeopteryx* then, helicopters now, and apoptosis forever.

John Funder

Baker Medical Research Institute,
PO Box 348, Prahran 3181, Australia

- Godfrey, J. *Nature* 373, 100 (1995).
- Jost, A. in *Mammalian Reproduction* (eds Gibian, H. & Platz, E. J.) 4–32 (Springer Verlag, Berlin, 1970).
- Pergament, E., Fiddler, M., Cho, N., Johnson, D. & Holmgren, W. J. *Hum. Reprod.* 9, 1730–1732 (1994).
- Tiffin, G. J., Rieger, D., Betteridge, K. J., Yadav, B. R. & King, W. J. *J. reprod. Fert.* 93, 125–132 (1991).
- Lentner, C. (ed.) *Geigy Scientific Tables* Vol. 1, 8th edition (Ciba-Geigy, Basle, Switzerland, 1981).
- Clarke, C. A. & Mittwoch, U. *Persp. biol. Med.* 37, 327–336 (1994).

The rise of neurogenetic determinism

Steven Rose

Dramatic advances in neuroscience are changing and enriching our understanding of brain and behaviour. But reductionist interpretations of these advances can cause great harm.

THIS 'Decade of the Brain' has produced dramatic advances in knowledge, together with ever more strident claims that neuroscience is about to 'solve' the brain and in doing so usher in a new era of what Delgado, an earlier enthusiast for brain surgery as a cure for violence, once called a "psychocivilised society"¹. The emerging synthesis of neurogenetics offers the prospect of identifying, ascribing causal power to and — if appropriate — modifying genes affecting brain and behaviour. Neurogenetics claims to be able to answer the question of where, in a world full of individual pain and social disorder, we should look to explain and to change our condition. If the faults lie outside ourselves, it is for the social sciences to understand and for politics to try to resolve the problems. But if the causes of our pleasures and our pains, our virtues and our vices, lie predominantly within the biological realm, then it is to neurogenetics that we should look for explanation, and to pharmacology and molecular engineering that we should turn for solutions.

Of course this is to simplify, to imply that the world is divided into mutually incommensurable realms of causation in which problems are *either* social *or* biological. This is not my intention; the phenomena of human existence and experience are always simultaneously biological and social, and an adequate explanation must involve both. Clearly, for any serious scientist to deny this would be equivalent to a politician denying that he or she was acting in the national interest; we are all interactionists now. But it is by their deeds that one must judge them, and in any search for explanation and intervention it is necessary to seek the appropriate level that effectively determines outcomes. Although only the most extreme reductionist would suggest that we should seek the origins of the Bosnian war in deficiencies in serotonin-reuptake mechanisms in Dr Karadzic's brain, and its cure by the mass prescription of Prozac, many of the arguments offered by neurogenetic determinism are not far removed from such extremes. Give the social its due, the claim runs, but in the last analysis the determinants are surely biological. And anyhow, we have some understanding and possibility of intervention into the biological, but rather little into the social.

This is not a new debate; it has recurred in each generation at least since Darwin's day, and most recently in the form of the

polemical disputes over the explanatory powers of sociobiology in the 1970s and 1980s (refs 2–5). What is new is the way in which the mystique of the new genetics is seen as strengthening the reductionist argument. At its simplest, neurogenetic determinism argues that there is a directly causal relationship between gene and behaviour. A man is homosexual because he has a "gay brain"⁶, itself the product of "gay genes"⁷, and a woman is depressed because she has genes "for" depression⁸. There is violence on the streets because people have "violent" or "criminal" genes⁹; people get drunk because they have genes "for" alcoholism¹⁰. In a social and political environment conducive to such claims, and which has largely despaired of finding social solutions to social problems, these apparently scientific assertions become magnified by press and politicians, and researchers may argue that their more modest claims are traduced beyond their intentions. Such Pilatism, however, is hard to credit when so much effort is put by researchers themselves into what Nelkin has described as "selling science"¹¹. The press releases surrounding the publication of LeVay's book⁶ or Hamer's paper⁷ were couched in language that leaves little need for media magnification.

It is my argument that such naive neurogenetic determinism is based on a faulty reductive sequence whose steps include: reification, arbitrary agglomeration, improper quantification, belief in statistical 'normality', spurious localization, misplaced causality, and dichotomous partitioning between genetic and environmental causes. The core issue is reducibility, which, as Medawar once remarked, comes not as second but as first nature to natural scientists. Thus when Popper, giving the first Medawar lecture to the Royal Society a few years back, offered eight terse reasons why biology was irreducible to physics (of which the fourth was that "biochemistry cannot be reduced to chemistry"), he incurred the wrath of the distinguished assembly¹². It is however not necessary to enter into a full-blown defence of irreducibility to identify the flaws in the claims of neurogenetic determinism to explain complex social phenomena. Such social phenomena are of their essence historically contingent and framed by meanings that the reductive process loses as surely as the information content of a page of *Nature* is absent

from a chemical study of the paper and ink that comprise it. And the issue at stake is not the formal philosophical one, but the question of the appropriate level of organization of matter at which to seek causally effective determinants of the behaviour of individuals and societies.

Reification converts a dynamic process into a static phenomenon. Thus violence, rather than describing an action or activity between persons, or even a person and the natural world, becomes instead a 'character' — aggression — a thing that can be abstracted from the dynamically interactive system in which it appears and studied in isolation. The same process occurs with 'intelligence', 'altruism', 'homosexuality' and so on. Yet if the activity described by the term violence can be expressed only in an interaction between individuals, to reify the process is to lose its meaning.

Arbitrary agglomeration carries reification a step further, lumping together many different reified interactions as all exemplars of the one thing. Thus aggression becomes the term used to describe processes as disparate as a man abusing his partner or child, fights between football fans, strikers resisting police, racist attacks on ethnic minorities, or wars. All are assumed to be manifestations of some unitary underlying property of the individuals, so that identical biological mechanisms are involved in, or even cause, each. Take for example the descriptions offered in a recent widely cited paper associating a point mutation in the structural gene for monamine oxidase inhibitor with "abnormal behaviour"¹³. The "behavioural phenotype" in eight males in this family is described as including "aggressive outbursts, arson, attempted rape and exhibitionism", activities carried out by subjects "living in different parts of the country at different times", across three generations. Can such widely differing types of behaviour, described so baldly as to isolate them from social context, appropriately be subsumed under the single heading of aggression? It is unlikely that such an assertion, if made in the context of a study of nonhuman animal behaviour, would pass muster. Yet the evidence it provides becomes part of the arsenal of argument employed, for example, by the current Federal Violence Initiative, which originated in 1992 in a proposal by the then director of the National Institutes for Mental Health, Frederick Goodwin, to

identify at least 100,000 inner-city children whose alleged biochemical and genetic defects will make them prone to violence in later life.

Improper quantification argues that reified and agglomerated characters can be given numerical value. If a person is violent, or intelligent, one can ask how violent or how intelligent by comparison with other people. IQ is one well-known example, but the quantification of aggression is also revealing, for it illustrates another feature of the reductionist cascade that leads to neurogenetic determinism, an animal model. Place an unfamiliar mouse into a cage occupied by a rat, and often the rat will eventually kill the mouse. The time taken for the rat to perform this act is taken as a surrogate for the rat's aggression; some rats will kill quickly, others slowly or even not at all. The rat that kills in 30 seconds is on this scale twice as aggressive as the rat that takes a minute. Such a measure, dignified as muricidal behaviour, serves as a quantitative index for the study of aggression, ignoring the many other aspects of the rat-mouse interaction, for instance the dimensions, shape and degree of familiarity of the cage environment to the participants in the muricidal interaction, whether there are opportunities for retreat or escape, and the history of interactions between the pair. And just as time to kill becomes a surrogate for a measure of aggression, so this behaviour in the rat is transmogrified into drive-by gangs shooting up a district in Los Angeles.

Belief in statistical normality assumes that in any given population the distribution of such behavioural scores takes a Gaussian form, the bell-shaped curve. The best known example is IQ, the tests for which successive generations of psychometricians refined and remoulded until it was made to fit (almost) the approved statistical distribution. But the assumption that the entire population can be distributed along a single dimension is to confuse a statistical manipulation for a biological phenomenon. There is no biological necessity for such a unidimensional distribution (even for continuously varying genetic traits), or for one in which the population shows such a convenient spread. (It is perfectly possible to set examinations so that virtually everyone scores 100 per cent — the UK university penchant for 10 per cent firsts, 10 per cent thirds and 10 per cent fails, with everyone else comfortably in the middle, is a convention, not a law of nature.) Yet the power of this reified statistic should not be underestimated. It conveniently conflates two different concepts of 'normality', implying that to lie outside the permitted

range around the norm is to be in some way abnormal. When Herrnstein and Murray called their new book *The Bell Curve*¹⁴ they played precisely into these multiple meanings of reified normality.

Having reified processes into things and arbitrarily quantified them, the reified object ceases to be a property even of the individual, but instead becomes that of a part of the person. Hence the penchant for speaking of, for example, schizophrenic brains, genes — or even urine — rather than of brains, genes or urine

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

"If homosexuality is inherited, shouldn't it have died out by now?"

derived from a person diagnosed as suffering from schizophrenia¹⁵. Of course, everyone ought to know (and does, at least on Sundays) that this is a shorthand, but the resonance of 'gay brains' or 'gay genes' does more than merely sell books for their scientific authors¹⁶; it both reflects and endorses the modes of thought and explanation that constitute neurogenetic determinism, for it disarticulates the complex properties of individuals into isolated and localized lumps of biology, permitting neuroanatomical debates to range over whether gayness is embedded in one or other hypothalamic region or, alternatively, a differently shaped corpus callosum^{5,17}. Aggression is 'located' in the limbic system, probably the amygdala. In the 1970s, US psychosurgeons proposed to treat inner-city violence by amygdectomy of ghetto militants¹⁸. Things are a little more sophisticated today; a localization in the brain can also take the form of some chemical imbalance, probably of neurotransmitters, so aggression is now 'caused' by a disorder of serotonineuptake mechanisms, and drugs rather than the knife become the approved approach¹⁹.

It is at this point that neurogenetic determinism introduces its misplaced sense of causality. It is of course probable that during aggressive encounters people show dramatic changes in, for instance, hormones, neurotransmitters and neurophysiological responses, all of which can be

affected by drug treatments. People whose life history includes many such encounters are likely to show lasting differences in a variety of brain and body markers. But to describe such changes as if they were the causes of particular behaviours is to mistake correlation or even consequence for cause. This issue has dogged the interpretation of the biochemical and brain correlates of psychiatric disorders for decades, yet it still continues; see for instance the recent debate over the role of dopamine D2 receptors in the aetiology of substance

abuse^{20,21}. It is however an almost inevitable consequence of the processes of reification and agglomeration, for if there is one single thing called alcoholism, for instance, then it becomes appropriate to seek a single causative agent; complexity is hard to deal with within the neurogenetic agenda.

Several single-gene defects lead to drastic dysfunctions of mind and brain. Huntington's disease, with its seemingly inevitable progress towards neurological collapse in middle age, is the classic example, but there are of course many others — Lesch-Nyhan syndrome, Tay-Sachs disease; the list of rare but devastating conditions is long. But neurogenetics consistently overstates its case, moving seam-

lessly from single to many genes, from genes with predictable consequences in virtually all known environments to genes with small or highly variable effects, whose norm of reaction extends so far as prevent any claims to predictability. In only a few per cent of all Alzheimer's cases is there a clear genetic involvement, whereas evidence identifying gene markers for manic depression and schizophrenia have been advanced with extensive publicity, and then quietly withdrawn²². At best, the hunt for the genes 'for' these conditions may be able to identify anomalous cases in which the genetic effect is to mimic a more widespread phenotypic condition (geneticists speak of phenocopies to emphasize the primacy of their genetic explanations in such cases; I have proposed the term 'genocopy' to help genetics appreciate its more limited contribution). What both concepts emphasize is the extent to which several pathways may lead to a final common biochemical or behavioural endpoint. What both mask is the possibility that the endpoints may not be in every sense identical. Some diagnosed depressions are ameliorated by one type of drug, some by another, and these distinctions have even been made the basis of diagnosis, in which the pharmacological response rather than the clinical syndrome is made the basis for a nosology (what Bignami has called *ex juvantibus* logic²³), which once again insists on the primacy of the biochemical over the behavioural, the biological over the social.

Drawing by Booth; © 1993/The New Yorker Magazine, Inc.

There are four main negative consequences of such determinism. The first is limited to the study of biology itself — the damage it does to conceptualizing living processes. The primacy given to 'genetic causes' fosters, even among researchers whose day-to-day practice ought to convince them otherwise, a linear view of living processes, in which the key to life itself lies in the one-dimensional string of nucleotide bases of DNA, the mythopoeic genome. Witness the metaphors with which molecular geneticists speak of the goals of the Human Genome Project — the 'Holy Grail', the 'Code of Codes', 'The Book of Life', 'genes for' particular conditions, as if the entire four dimensions of an organism (three of space and one of developmental and life history) can be read off from this linear code, like a telephone directory.

The shorthand phrase of a gene 'for' a condition is profoundly misleading — after all, there aren't really even genes 'for' blue or brown eyes, let alone such complex and historically and socially shaped features of human existence as sexual desire or urban guerillas. The cellular developmental and enzymatic route that results in the manufacture of particular pigments involves many thousand genes; the route that leads to the behavioural manifestations we call desire clearly involves genes, but cannot sensibly be regarded as abstractly embodied in them. What there are, of course, are differences between genomes, genes in the absence of which differences in eye colour or other phenotypes may emerge. The biologist looking at the effects of particular genetic mutations or deletions studies the functioning of the system in the absence or malfunction of a particular gene. Furthermore, the system is not a passive responder to absence or malfunction, but seeks, by means of developmental plasticity, to compensate for any deficit — a point sometimes lost on those hoping to draw clear-cut conclusions from the often rather modest behavioural consequences of knock-out experiments where mice in which genes coding for apparently essential proteins have been deleted.

A considerable disservice was done to biology by the historical chance that meant that in the early years of this century two separate subdisciplines emerged — genetics, which essentially asks questions about the origins of differences between organisms; and developmental biology, which asks questions about the processes that ensure similarity. The careless language of DNA and molecular genetics serves to widen this gap rather than to help bridge it and thereby open the route towards the synthetic biology that we so badly need. As is well known, chimpanzees and humans share upwards of 98 per cent of their DNA, yet no one would confuse the two phenotypes. We have no

idea at present about the developmental rules that lead in one case to the chimp, in the other to the human. But this, surely one of the great unsolved riddles of biology, seems a matter of indifference to most molecularly oriented geneticists.

The other consequences of determinism reach out beyond theory. If the homeless or depressed are so because of a flaw in their biology, their condition cannot be the fault of society, albeit a humane society will attempt, pharmacologically or otherwise, to alleviate their distress. This 'victim blaming' generates in its turn a sort of fatalism among those it stigmatizes; it is not our fault, the problem lies in our biology. Such fatalism can bring its own relief, for less stigma attaches to being the carrier or transmitter of deficient genes than to having been morally responsible. It is striking that in the United States, leading gay activists have embraced the gay brains/gay genes explanation for their sexual orientation on the explicit grounds that they can no longer be held morally culpable for a 'natural' state, nor can they be seen as dangerously likely to infect others with their 'perverse' tastes.

The final consequence is the subversion of scarce resources. Support for research and treatment becomes misfocused. The orientation of research funds in Russia towards the molecular biology and genetics of alcoholism is one good example (for all that any rational attempt to explain the prevalence of vodka-sodden drunks on the streets of Moscow would not immediately seize on the peculiar genetics of the Russian population as its starting point). Similarly, the Violence Initiative, directed towards seeking the origins of violence in US society in terms of the genotypes of blacks and poor inner-city whites, the problems of 'temperament' in toddlers and the deficiencies in serotonin-reuptake mechanisms in incarcerated criminals, is clearly going to keep a generation of psychologists, neuropharmacologists and behavioural geneticists in research funds for a good few years to come²⁴. But as an approach to diminishing the violence of city streets it would seem unlikely to achieve as significant an impact as would measures to reduce the estimated 280 million handguns currently in personal possession in the United States.

Even in less dramatic instances, emphasis on genetic explanations and molecularly oriented research prevents researchers from seeing and studying the obvious. The almost universal conviction among biological psychiatrists that schizophrenia is a genetic disorder means that they are unable to respond to the suggestive epidemiological evidence that the diagnosis of schizophrenia in the children of black-white relationships in the United Kingdom is several times higher than that of either of the parental populations²⁵. No genetic model fits this finding as well as an

explanation in terms of the racism of the society in which these children grow up. Yet it is well known that with a little ingenuity any phenotypic distribution can be explained genetically granted appropriate assumptions about partial penetrance and incomplete dominance, and it is not hard for a behavioural geneticist to offer as an alternative that the data could be accounted for by assortative mating (E. Gershon, personal communication).

There is no doubt that the dramatic increases in neuroscientific knowledge are changing and enriching our understanding of brain and behaviour. There is equally no doubt that, wisely and appropriately employed, the new knowledge offers the potential to diminish the degree of human suffering, at least in relatively wealthy industrialized societies. But the new findings must be broken out of their reductionist mould and relocated within a more integrated understanding of the relationships between the biological, personal and social; we must abandon the unidirectional view of the causes of human action so as to recognize the appropriate, determining level of explanation for complex phenomena. Otherwise the findings' potential for good remains limited and for misapplication substantial and disturbing. □

Steven Rose is in the Department of Biology, The Open University, Milton Keynes MK7 6AA, UK.

1. Delgado, J. M. R. *Physical Control of the Mind: Towards a Psychocivilised Society* (Harper & Row, New York, 1971).
2. Wilson, E. O. *Sociobiology: The New Synthesis* (Harvard Univ. Press, Cambridge, MA, 1975).
3. Dawkins, R. *The Selfish Gene* (Oxford Univ. Press, Oxford, 1976).
4. Rose, S. P. R., Lewontin, R. C. & Kamin, L. J. *Not in Our Genes* (Penguin, Harmondsworth, 1984).
5. Kitcher, P. *Vaulting Ambition: Sociobiology and the Quest for Human Nature* (MIT Press, Cambridge, MA, 1985).
6. LeVay, S. *The Sexual Brain* (MIT Press, Cambridge, MA, 1993).
7. Hamer, D., Hu, S., Magnuson, V. L., Hu, N. & Pattatucci, A. M. L. *Science* **261**, 321–327 (1993).
8. Cohen, D. B. *Out of the Blue: Depression and Human Nature* (Norton, New York, 1994).
9. Reiss, A. & Roth, J. *Understanding and Preventing Violence* (National Academy Press, New York, 1993).
10. Galanter, M. (ed.) *Recent Developments in Alcoholism* (Plenum, New York, 1993).
11. Nelkin, D. *Selling Science* (W. H. Freeman, New York, 1987).
12. Perutz, M. *Trends biochem. Sci.* **13**, 206 (1988).
13. Brunner, H. J., Nelen, M., Breakefield, X. O., Ropers, H. H. & van Oost, B. A. *Science* **262**, 578–580 (1993).
14. Herrnstein, R. J. & Murray, C. *The Bell Curve: Intelligence and Class Structure in American Life* (Free Press, New York, 1994).
15. Rose, S. P. R. *Trends Neurosci.* **10**, 152 (1987).
16. Hamer, D. & Copeland, P. *The Science of Desire: The Search for the Gay Gene and the Biology of Behavior* (Simon & Schuster, New York, 1994).
17. Swaab, D. F. *Eur. J. Neurosci. Suppl.* **7**, 126 (1994).
18. Mark, V. H. & Ervin, F. R. *Violence and the Brain* (Harper & Row, New York, 1970).
19. Kramer, P. *Listening to Prozac* (Viking Penguin, New York, 1994).
20. Uhl, G., Blum, K., Noble, E. & Smith, S. *Trends Neurosci.* **16**, 83–88 (1993).
21. Chipkin, R. E. *Trends Neurosci.* **17**, 50 (1994).
22. Mann, C. C. *Science* **264**, 1686–1689 (1994).
23. Bignami, G. in *Against Biological Determinism* (ed. Rose, S. P. R.) 94–110 (Allison & Busby, London, 1982).
24. Breggin, P. R. & Breggin, G. R. *Int. J. Psychol. Psychother.* **11**, 59–71 (1993).
25. Harrison, G. *Schiz. Bull.* **16**, 663–671 (1990).

The endless search for accelerators

Particle physics is far from dead, nor is the search for new ways of accelerating particles to high energy. A new demonstration shows that there are still important things to find. Who will say that nothing will come of them?

WITH the Superconducting Super Collider (SSC) just a half-dug tunnel in Texas, it may seem strange, to the world outside particle physics, that people are still working up novel schemes for the acceleration of particles. But hope springs eternal. Suppose that, with a little ingenuity, 10^{12} electron-volt accelerators become affordable? Will not the whole world beat a path to their inventors' door? And in any case, who would wish to suppress the exercise of ingenuity?

In principle, particle accelerators have not much changed since the invention of the cyclotron in 1930. You start with charged particles essentially at rest and accelerate them by the application of external electrical and magnetic forces, the former as a source of energy, the latter for confinement. Of course, in more than six decades, there have been great advances in sophistication as well as in size. Who would have thought, for example, that it would now be standard practice, at roughly circular machines, for the control of circulating bunches of particles to be done by light signals sent along a diameter?

A cyclotron has a fixed magnetic field, but a synchrotron (such as SSC would have been) is a machine in which the magnetic field increases with the growing momentum of the accelerated particles. But the mechanism is, in its barest essentials, unchanged. And the outwardly different linear accelerators are in principle the same. In no way does this belittle the cleverness of these machines (and of their designers). And costs are broadly a function of the weight of iron (or its equivalent as superconducting niobium).

That is the spur to ingenuity among those who would accelerate particles — the hope that a radically different method would be radically less expensive. These days, of course, with high-power lasers all the rage, light as a means of acceleration is an obvious place to look for a radical solution. That line of argument seems to have been the inspiration for a remarkable demonstration at the Brookhaven National Laboratory that Čerenkov radiation or, more accurately, its inverse, might do the trick.

Čerenkov radiation is a phenomenon whose unfamiliarity has long since peaked. Twenty years ago, students were invited to marvel at what happens when a relativistic charged particle moves through a refractive material with a velocity exceeding that of the allowable velocity of light (reduced from the velocity of light *in vacuo* by division by the local refractive index). What happens is

that a conical sheet of radiation, tightly wrapped around the forward direction of motion, accompanies the particle. With care, the pulses of radiation can be detected at fixed photon detectors and, with a little triangulation, used to define the movement of charged particles through the atmosphere. Now, there are commercial instruments working on this principle, but the marveling is somewhat faded.

The Brookhaven demonstration would make the Čerenkov effect work backwards. A charged particle moving through a medium will emit a cone of radiation divergent in the forward direction, right? Then why not illuminate such a particle moving through a medium with an appropriate conical pulse of radiation converging on the forward direction of motion? It needs no formal solution of Maxwell's equations, but only a little handwaving about symmetry, to show that the result should be an acceleration.

The demonstration has been mounted by a group of nine people drawn from Brookhaven itself, from STI Optronics at Bellevue, Washington State, and a lone physicist (Y. Liu) from the University of California at Los Angeles (Kimura, W.D. *et al. Phys. Rev. Lett.* **74**, 546-549; 1995). It is an elaborate demonstration and, at this stage, not especially conclusive, as will be learned.

But Kimura *et al.* come away from their work with a striking talisman; there is a sense in which they have been able to accelerate electrons by 3.7 MeV in a 12-cm interaction space, corresponding to some 30 MeV per metre. Packing industrial or medical accelerators into suitcases is plainly in somebody's mind. But what the world needs to make particle physics affordable are machines that add something like 1 GeV per metre to the energy of a particle. There is a factor not far below 10^5 to be gained before the dream of the 100-metre linear equivalent to the SSC can be realized. But one must start somewhere, after all.

And this is a demonstration only. This is how it works. The working fluid, so to speak, is a 40-MeV electron beam and the driving force is a 10-GW CO_2 pulsed laser (not every laboratory will have one). The crucial element of the arrangement is a device called an axicon, which is a narrowly convergent convex mirror that also rotates the local plane of polarization of the linearly polarized laser beam so as to create the illusion that the electromagnetic oscillations are coherently radial across the whole field.

The experiment is done in hydrogen gas; the electron beam enters and leaves the gas chamber through diamond windows a few micrometres thick and one mm or so in diameter. The Čerenkov angle (half the angle subtended by the beam) amounts to 20 mrad and thus, at about 1° , is a powerful testament to the axicon-maker's art (which appears to be mostly STI Optronics).

The energy of the electrons is measured with an energy-spectrometer which, at this stage, appears to have too small a dynamic range for convenience. But what does emerge is that at least some fraction of the electrons in the beam are substantially accelerated, by an energy up to 3.7 MeV. It is also clear, from the data presented, that some electrons are decelerated. (Phase matching is a crucial business in the Čerenkov business, but the electrons in the authors' electron beam are as likely to be out of phase as in it.)

Where is all this likely to lead? The authors' back-of-the-envelope calculations centre on the laser power. If the energy output could be increased to 5 GW at its peak, and the length of the interaction region increased by a factor of 2, the energy gradient would be 60 MeV per metre. And a 100-GW laser would give a gradient of some hundreds of MeV per metre. But that, sadly, is still a factor of 10,000 away from the 1 GeV per metre accelerator.

So Brookhaven will not be able to replace the SSC on the strength of the present performance of its inverse-Čerenkov accelerator. Does that mean that the demonstration simply shows that the accelerator community should be looking somewhere else?

Nothing could be further from the truth. The expectation that the Čerenkov effect is reversible may leap from the pages of the textbooks, but it has not previously been confirmed on such a scale. Moreover, the heavy laboratory engineering that seems to have been unavoidable in this demonstration is incommensurate with the scale on which future particle accelerators are likely to operate in the distant future. What particle physics needs for the future, after all, is not a great flux of accelerated particles but individual particles guided with the precision required to ensure that most of those given high energy can be used. That argues for a scaling down, not a scaling up, of the Brookhaven demonstration. Nobody should be surprised if that is called inverse Compton scattering.

John Maddox

Causes of insulin resistance

C. Ronald Kahn

INSULIN resistance is a fundamental component of the pathogenesis of non-insulin-dependent diabetes mellitus (NIDDM), also known as type II diabetes, and is associated with a variety of other disorders such as obesity, hypertension, polycystic ovarian disease and atherosclerosis. But for all the importance of this condition, its molecular basis remains poorly understood¹. On page 448 of this issue, Maddux *et al.*² present evidence for a hitherto unknown mechanism for insulin resistance in some patients with NIDDM. In these cases, resistance is related to overexpression of a plasma membrane glycoprotein called PC-1, which the authors identified by purifying and sequencing fractions which inhibit insulin-receptor kinase activity.

At the cellular level, insulin action involves a complex network of molecules^{3,4} (see figure). Following insulin binding to the α -subunit of its receptor, the tyrosine kinase in the β -subunit undergoes autophosphorylation and activation. This results in phosphorylation of several intracellular substrates on tyrosine residues, the best characterized of which is insulin receptor substrate-1 (IRS-1). This is a cytoplasmic protein with multiple tyrosine phosphorylation sites which, following insulin stimulation, serve as docking sites for intracellular molecules that contain specific recognition domains, termed SH2 domains. Some of the SH2 domain-containing molecules are enzymes, such as phosphatidylinositol-3-OH kinase (PI3K), and are directly activated by docking; others are adaptors which link insulin signalling to the Ras pathway and families of serine-threonine kinases termed the raf, MAP and S6 kinases.

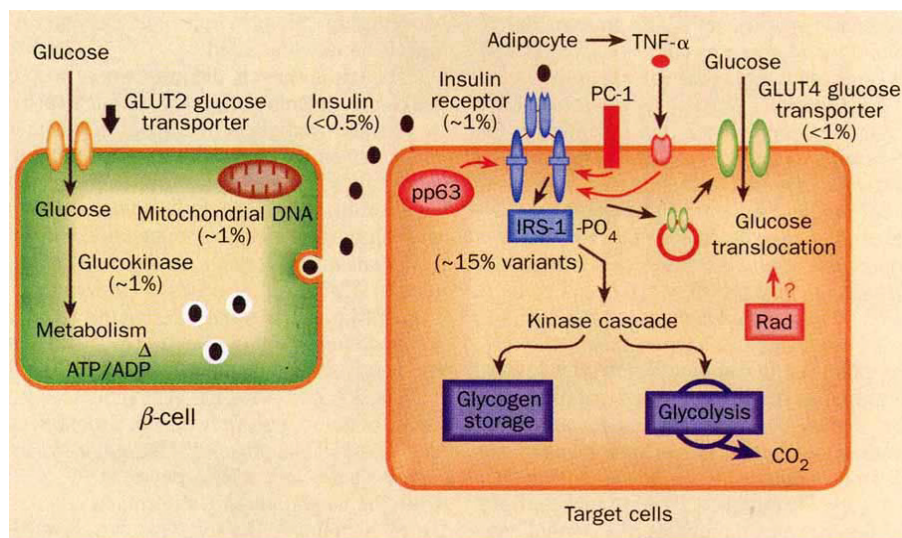
Ultimately, this process leads to stimulation of glycogen, lipid and protein synthesis, as well as translocation of insulin-sensitive glucose transporters to the surface of muscle and fat cells. Insulin-resistant states, such as NIDDM, are characterized by a decrease in the ability of insulin to stimulate glucose uptake into muscle and fat, as well as to inhibit glucose production by the liver.

Attempts to define the exact locus of insulin resistance in NIDDM have met with limited success. By the time clinical diabetes is manifest, there is a decrease in insulin-receptor number and tyrosine-kinase activity, and defects in glucose transporter translocation, MAP and S6 kinase activation, and glycogen synthesis^{5,6}. Although not all of the molecules involved in insulin action have been studied, available data have not pointed to any of the primary candidate genes as a major locus of mutation in NIDDM.

Genetic defects in the insulin receptor have been found in more than 40 people with diabetes, but most of them are patients with rare syndromes of extreme insulin resistance^{7,8}. About 15 per cent of patients with NIDDM (and 5 per cent of non-diabetics) have genetic variation in the sequence of IRS-1 (ref. 9), though the functional significance of this is unclear because transgenic mice lacking IRS-1 use alternative pathways of insulin action and have only mild glucose intolerance^{10,11}. Genetic defects in other components of

By contrast, Maddux *et al.*² now demonstrate that the membrane glycoprotein PC-1 not only inhibits the insulin receptor kinase, but that its levels are higher than normal in fibroblasts and skeletal muscle of some patients with NIDDM. Furthermore they show that when breast carcinoma cells are transfected with PC-1, there is a decrease in insulin-stimulated mitogenesis. Although PC-1 has several different enzymatic activities, the exact mechanism by which it acts as an inhibitor of the insulin-receptor kinase is unknown. It is also unclear whether PC-1 inhibits other insulin-sensitive pathways, such as stimulation of glucose uptake and glycogen storage.

Two other inhibitors of insulin action



Non-insulin-dependent diabetes mellitus results from a combination of insulin resistance at the target cell (liver, muscle and fat) and relative deficiency in insulin production from the β -cell. Genetic defects may alter β -cell or target cell function in a small percentage of patients (numbers in parentheses indicate the approximate frequency of genetic mutation in NIDDM). Down-regulation of GLUT2 glucose transporters in β -cells and of insulin receptors in target cells also contributes to the pathogenesis of the disease. Potential inhibitors of insulin action in target cells are shown in red; they include PC-1, pp63 (α -2-HS-glycoprotein), a Ras-related protein termed Rad, and signals mediated by the TNF- α receptor. The PC-1-, pp63- and TNF-mediated signals are all believed to act directly to reduce insulin-receptor tyrosine kinase activity, but the exact mechanisms remain unknown. Rad is thought to act intracellularly to alter GLUT4 glucose transporter translocation to the plasma membrane.

the insulin-action pathway seem to be extremely rare, which has prompted several groups to try to identify molecules that might serve as diabetogenes by acting as inhibitors of insulin action.

Although there has been indirect evidence of such inhibitors for some time, the first one to be purified and cloned was a rat liver phosphoprotein known as pp63 (ref. 12) and subsequently identified as α -2-HS-glycoprotein¹³. This protein acts as an inhibitor of the insulin-receptor tyrosine kinase and inhibits insulin-stimulated IRS-1 phosphorylation, PI3K activation and insulin-stimulated DNA synthesis, without affecting amino-acid uptake or tyrosine aminotransferase activity. But so far there has been little evidence that α -2-HS-glycoprotein is a direct inhibitor of insulin action in NIDDM.

also need consideration in NIDDM. One is a member of the Ras/GTPase superfamily, termed Rad, expression of which is increased in skeletal muscle of some patients with NIDDM¹⁴. Rad is believed to act as an intracellular inhibitor of insulin action, possibly at the level of glucose-transporter translocation. Circulating inhibitors of the insulin receptor have also been postulated to play a role in NIDDM. The most interesting of these is tumour necrosis factor- α (TNF- α), which is produced in excess by adipocytes of obese individuals and results in a decrease in insulin-receptor kinase activity in muscle and fat¹⁵.

Assuming intact feedback loops, these defects in insulin action should be compensated for by an increase in insulin secretion. This does not occur in people

with NIDDM, however, because the β -cells are desensitized to glucose (another defect for which the mechanism remains poorly understood). So although significant progress has been made in defining the molecular mechanisms of insulin action and alterations in rare syndromes of insulin resistance, we still do not know the molecular and genetic mechanisms underlying the fundamental pathogenesis in

most individuals with NIDDM. Further investigation will be required to identify the main sites of insulin resistance, and to determine the true position of PC-1 in the list of causes of this disorder. □

C. Ronald Kahn is at Harvard Medical School and the Joslin Diabetes Center, One Joslin Place, Boston, Massachusetts 02215, USA.

1. Kahn, C. R. *Diabetes* **43**, 1066–1084 (1994).
2. Maddux, B. A. *et al. Nature* **373**, 448–451 (1995).
3. White, M. F. & Kahn, C. R. *J. biol. Chem.* **269**, 1–5 (1994).
4. Avruch, J. *et al. FASEB J.* **41**, 2629–2633 (1993).
5. Haring, H. U. & Mehnert, H. *Diabetologia* **36**, 176–182 (1993).
6. Sommercorn, J., Fields, R., Raz, I. & Maeda, R. *J. clin. Invest.* **91**, 509–513 (1993).
7. Moller, D. A. & Flier, J. S. *New Engl. J. Med.* **325**, 938–948 (1991).
8. Taylor, S. I. *et al. Endocrinol. Rev.* **3**, 566–595 (1992).
9. Almind, K. *et al. Lancet* **342**, 828–832 (1993).
10. Araki, E. *et al. Nature* **372**, 186–190 (1994).
11. Tamemoto, H. *et al. Nature* **372**, 182–186 (1994).
12. Auberger, P. *et al. Cell* **58**, 631–640 (1989).
13. Srinivas, P. R. *et al. Molec. Endocrinol.* **7**, 1445–1455 (1993).
14. Reynet, C. & Kahn, C. R. *Science* **262**, 1441–1444 (1993).
15. Hotamisligil, G. S., Budavari, A., Murray, D. & Spiegelman, B. M. *J. clin. Invest.* **94**, 1543–1549 (1994).

APOPTOSIS

Death of a T cell

Andreas Strasser

WAS it suicide or murder, and how was the deed done? These are questions not only for the police but for immunologists trying to find out how T-lineage cells stimulated through their antigen receptors (TCR) meet their end. Three papers in this issue^{1–3} — on pages 438, 441 and 444 — provide evidence that activation of the cell-surface receptor known as Fas, APO-1 or CD95 is the weapon involved and that the process, at least as it happens *in vitro*, is suicidal.

In the thymus, deletion of immature T cells that happen to express self-antigen-reactive receptors is a safeguard against autoimmunity (central tolerance). In peripheral lymphoid tissues, the physiological death of TCR-activated, mature T cells probably serves to delete autoreactive T cells with specificity for antigens that are not presented in the thymus (peripheral tolerance), and to terminate an immune response after the battle against a pathogen has been won. Activated T cells are a potential hazard. They produce copious amounts of inflammatory cytokines, and contribute to infection-induced hyperplasia in lymph nodes (potentially a forerunner of lymphoid malignancy); it is therefore safer to remove them when their job has been done and they are no longer needed.

Activation-induced death was first observed in T-cell hybridomas⁴. These cells undergo growth arrest, functional activation — for example secretion of interleukin-2 (IL-2) — and apoptosis upon stimulation with antigens, mitogens or antibodies specific for TCR/CD3, Thy-1 or Ly-6. Signal transduction initiated by all these stimuli requires a functional TCR/CD3 complex and *de novo* macro-

molecular synthesis. Two transcription factors, the orphan steroid receptor Nur77 (refs 5, 6) and Myc-Max heterodimers⁷, are essential for TCR-activation-induced apoptosis but are not necessary to induce IL-2 secretion, indicating that these two processes are subject to distinct genetic control. Death induced by TCR activation also occurs *in vitro* in normal and leukaemic T-cell lines, and in normal T lymphoblasts¹.

The CD95 receptor and its ligand were implicated in initiating this form of apoptosis by Russell *et al.*⁸, who showed that mature, activated T cells from mutant *lpr* and *gld* mice are resistant to apoptosis induced by reactivation through the TCR. Mutant *lpr* mice are deficient in functional expression of CD95 (ref. 9), a member of the tumour-necrosis factor (TNF) receptor family; mutant *gld* mice are likewise deficient in CD95 ligand¹⁰, a member of the corresponding family of TNF-related cytokines.

The new work reveals that autocrine stimulation of the CD95 receptor is essential for TCR-activation-induced death of T hybridoma cells^{2,3}, Jurkat T leukaemia cells¹ and non-transformed, preactivated T cells¹. All three papers show that TCR stimulation induces the expression of both CD95 and its ligand on these cells, and that killing can be inhibited by blocking either the receptor or the ligand. Blocking was achieved with non-stimulatory, CD95-specific antibody fragments or a chimaeric CD95-immunoglobulin fusion protein. Induced expression of CD95 ligand was inhibited by cyclosporin A (ref. 1), and CD95 was downregulated by glucocorticoids³, explaining why this route to death can be blocked by these

two agents. Experiments with individually cultured Jurkat and T hybridoma cells demonstrated formally that each TCR-activated cell produces all of the elements necessary for its destruction, and that death is prevented by blocking CD95 or its ligand^{1,2}.

These papers describe work carried out *in vitro*, however. T hybridoma cells have characteristics of both their parents, the mature, activated normal T cells (IL-2 secretion) and the immature, transformed thymic lymphoma cells (high sensitivity to killing by glucocorticoids and γ -irradiation). It is not known which parent bestowed upon T hybridoma cells their suicidal abilities, so it is unclear whether this system is a model for apoptosis of autoreactive thymocytes, activated mature T cells, or both. Experiments with cells from mutant *lpr* and *gld* mice have established that both CD95 and its ligand are essential for TCR-activation-induced apoptosis of peripheral, mature T cells^{8,11}, but not for deletion of autoreactive thymocytes¹¹. The unusual TCR α/β^+ , CD4⁺, CD8⁺, B220⁺ cells that accumulate in these mice are therefore probably not autoreactive escapees from intrathymic deletion, but descendants of previously activated, peripheral T cells that failed to undergo physiological death at the end of an immune response.

It is unclear whether death of activated, mature T cells under physiological circumstances is strictly suicidal, or whether CD95 ligand is effectively provided by neighbouring, activated T cells, or maybe even antigen-presenting cells. Normal T cells can repress the mutant *gld* phenotype in chimaeric mice¹², suggesting that suicide is not the only way for an activated, mature T cell to die. It is also unclear whether most CD95 ligand *in vivo* is membrane-anchored, as in T hybridoma cells², or whether functionally significant quantities are secreted, as from Jurkat cells¹. Some of these questions can now be addressed using soluble chimaeric CD95-immunoglobulin fusion proteins as ligand-specific probes.

Because TCR stimulation triggers not only proliferation and functional activation but also cell death, how can a T-cell-dependent immune response be generated? The answer may lie in the observations that although T cells upregulate CD95 expression within 24 hours of TCR stimulation, they only become sensitive to CD95-mediated apoptosis several days later^{13,14}. Maybe this phase is all the time that they have to carry out their immune functions. Indeed, CD95 triggering may even be beneficial during the initial phase of T-cell activation — CD95-specific antibodies can cooperate with suboptimal concentrations of TCR/CD3-specific antibodies in mitogenesis¹⁵, and T cells from mutant *lpr* mice are less responsive to antigenic stimulation than normal T

cells¹¹. The molecular mechanisms that determine sensitivity or resistance to CD95-ligand-induced apoptosis must be central to the proper functioning of the immune system.

Why does deletion of autoreactive thymocytes proceed normally in mutant *lpr* and *gld* mice; and why do these mice have detectable abnormalities only in B and T cells, although CD95 expression is widespread? It could be that CD95 and its ligand are strictly involved in the killing of activated, mature B and T cells, but I favour two alternative explanations. One is that *lpr* and *gld* mice are only 'knock-down' mutants which retain sufficient activity to kill autoreactive thymocytes and perhaps other cells. The other is that an additional, possibly related, receptor-ligand pair operates in some cells, such as thymocytes, but not in mature, activated T cells. Generation of mice in which CD95 and its ligand have been knocked out, and crossing them with mice deficient in other members of the TNF receptor or TNF cytokine families, should clarify these issues.

Although we still don't know whether, *in vivo*, death induced by TCR activation is suicide or murder, CD95-triggering has now been identified as the weapon. One of the next steps will be to find out what happens after CD95-triggering to see if these events are connected to events following cell death induced by other stimuli. Indeed, it has already been shown that apoptosis induced by TCR activation, like other routes to physiological death, depends on the activity of a serine or cysteine protease¹⁶. □

Andreas Strasser is at The Walter and Eliza Hall Institute of Medical Research, Post Office, The Royal Melbourne Hospital, Victoria 3050, Australia.

A rougher, tougher Moon

Sean C. Solomon

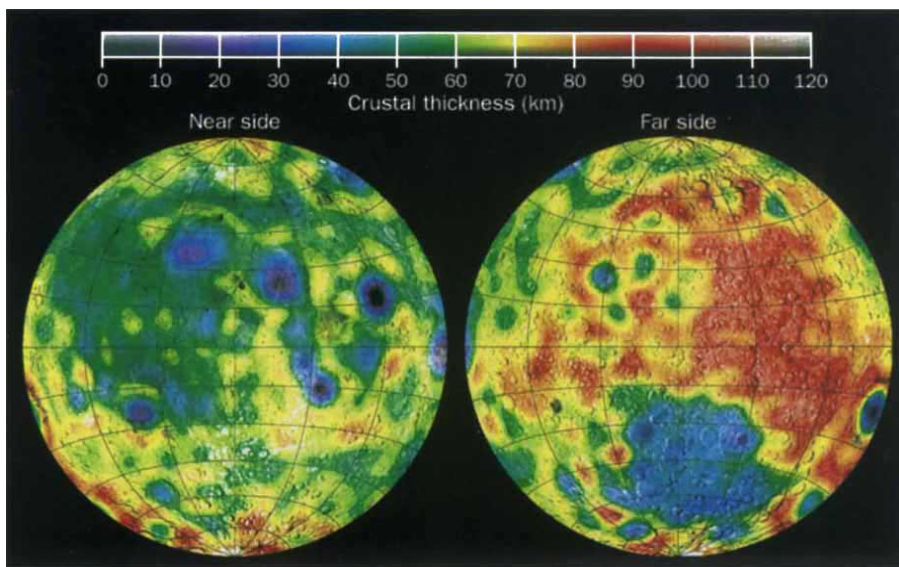
"THERE is nothing left remarkable beneath the visiting moon", lamented Shakespeare's Cleopatra on the death of Mark Anthony. Such a sentiment seemed also to guide the space programmes of the United States and Soviet Union in the years that followed the extraordinary burst of lunar exploration, encouraged by the political imperatives of the Cold War, that filled the 1960s and early 1970s. That complacent view, however, has been shattered by the first findings of the recent Clementine mission to the Moon, itself a beneficiary of Cold War technology originally developed for the 'Star Wars' anti-missile defence system, as a collection of papers now shows¹⁻⁴. The most interesting of the new results, from the first global topographic survey and an improved lunar gravity field, are changing our understanding of the early history of the Moon.

The prevailing view of lunar evolution^{5,6} following the Apollo and Luna missions was that the accretional energy of lunar formation melted much of the outer part of the earliest Moon, and that gravitational separation of crystals from liquid during cooling led to the formation of much of the crust. Over the first 700 to 800 million years of lunar history, the Moon was bombarded by a heavy shower of meteoroidal and cometary material, including dozens of objects so massive that their impacts formed multi-ring basins hundreds of kilometres

in diameter. During this era, the still-cooling crust was thought to be sufficiently deformable that a state of isostasy prevailed; that is, regions of higher elevation were underlain by a thicker layer of low-density crustal material such that at a given level beneath the crust the pressure of an overlying column of rock was roughly constant.

During and following the period of heavy bombardment, basaltic magma generated by partial melting of the mantle underlying the lunar crust erupted into low-lying regions to form the lunar maria. The maria occur preferentially on the lunar near side, a result attributed to the generally greater elevations and thicker crust of the Moon's far side⁷. During at least the later stages of mare volcanism, the Moon had developed a mechanically strong outer shell, or lithosphere, that resisted deformation despite the weight of the mare basalt deposits⁸. Evidence of this strength is preserved today in the strong positive gravity anomalies, or mascons⁹, seen over a number of the circular maria that fill large impact basins.

The Clementine spacecraft, in lunar orbit between February and May of last year, carried a laser ranging instrument that functioned as an altimeter² as well as four imaging systems¹ that operated in a number of distinct optical and infrared wavelength bands between 0.4 and 8 μm . About 70,000 individual measures of



Thickness of the Moon's crust (Lambert equal-area projection), obtained from the newly determined lunar topography and gravity fields under the assumptions that crust and mantle are each of uniform density and that the crustal thickness in the area bounded by the Apollo landing sites is that determined from the Apollo seismic experiment² (courtesy of M. T. Zuber and G. A. Neumann). The crustal thickness has been underestimated beneath the maria because of the neglect of the effect of the high-density mare basalt material on the gravity field. The South Pole-Aitken basin is underlain by the roughly circular region of thin crust centred in the lower half of the right-hand map.

1. Dhein, J., Walczak, H., Bäuml, C., Debatin, K.-M. & Krammer, P. H. *Nature* **373**, 438–441 (1995).
2. Brunner, T. *et al. Nature* **373**, 441–444 (1995).
3. Ju, S.-T. *et al. Nature* **373**, 444–448 (1995).
4. Ashwell, J. D., Cunningham, R. E., Noguchi, P. D. & Hernandez, D. J. *exp. Med.* **165**, 173–194 (1987).
5. Woronicz, J. D., Calnan, B., Ngo, V. & Winoto, A. *Nature* **367**, 277–281 (1994).
6. Liu, Z.-G., Smith, S. W., McLaughlin, K. A., Schwartz, L. M. & Osborne, B. A. *Nature* **367**, 281–284 (1994).
7. Bissonnette, R. P., McGahon, A., Mahboubi, A. & Green, D. R. *J. exp. Med.* **180**, 2413–2418 (1994).
8. Russell, J. H., Rush, B., Weaver, C. & Wang, R. *Proc. natn. Acad. Sci. U.S.A.* **90**, 4409–4413 (1993).
9. Watanabe-Fukunaga, R. *et al. Nature* **356**, 314–317 (1992).
10. Takahashi, T. *et al. Cell* **76**, 969–976 (1994).
11. Singer, G. G. & Abbas, A. K. *Immunity* **1**, 365–371 (1994).
12. MacDonald, G. C., Kakkanaiah, V. N., Sobel, E. S., Cohen, P. L. & Eisenberg, R. A. *J. Immun.* **154**, 444–449 (1995).
13. Owen-Schaub, L. B., Yonehara, S., Crump, W. L. & Grimm, E. A. *Cell. Immun.* **140**, 197–205 (1992).
14. Klas, C., Debatin, K.-M., Jonker, R. R. & Krammer, P. H. *Int. Immun.* **5**, 625–630 (1993).
15. Alderson, M. *et al. J. exp. Med.* **178**, 2231–2235 (1993).
16. Sarin, A., Adams, D. H. & Henkart, P. A. *J. exp. Med.* **178**, 1693–1700 (1993).

lunar elevation were used to construct a 2° by 2° grid of topography complete for latitudes less than 75° from the equator² (1° of latitude on the Moon is about 30 km). Information on the lunar gravity field was obtained from measurements of Doppler shifts in the frequency of the radio signal transmitted to Earth by the spacecraft, and the new observations were combined with data from earlier orbiters to produce a spherical harmonic representation of the gravity field² complete to harmonic degree and order 70 (corresponding to spatial wavelengths of about 150 km and greater).

The first surprise was a greater than expected topographic relief. Lunar elevation varies by 16 km, a figure more than 30 per cent greater than the relief known from earlier measurements. Both the topography and the gravity fields show clear anomalies over many impact basins, including several features only tentatively identified as basins before the Clementine data provided confirming evidence³. The most pronounced topographic feature on the Moon is the South Pole–Aitken basin⁵, which at 2,500 km diameter and 8 km deeper than the lunar reference surface is now championed as the largest impact structure in the Solar System^{2,3}. The formation of such a large basin by impact may have excavated material from the underlying mantle as well as from the crust. Multispectral imaging by Clementine has verified earlier suggestions that the surface of the South Pole–Aitken basin floor is richer in ferromagnesian minerals than is typical of lunar highland areas, consistent with an admixed component of mantle material^{3,4}.

Another surprise was that the gravity and topography of some impact basins largely unfilled by mare basalt deposits indicate departures from a fully isostatic state that resist easy interpretation. The Mendel–Rydberg basin, for instance, about 600 km in diameter and 5 km deep, has a modest gravity anomaly and is consistent with complete isostatic compensation of relief² by a crust left thinner than normal by the collapse of the impact-excavated cavity and subsequent crustal flow. The nearby South Pole–Aitken basin, however, the oldest⁵ as well as the largest recognized basin on the Moon, seems to be only about 90 per cent compensated². An isostatic state beneath a basin requires that the underlying crustal rocks be sufficiently hot to flow on time-scales of a few hundreds of millions of years or less. The variable state of isostasy beneath lunar basins indicates that such conditions were not simply a function of either the age of the cooling Moon or the energy of the impact event (which scales as a power of the basin size). Rather the strength, and presumably the temperature structure, of the Moon's lithosphere dur-

ZOOLOGY

In the eye of the kestrel

WINGED raptors — such as this kestrel — are renowned for their keen eyesight. But unlike humans, kestrels can also see in the ultraviolet, and in the report on page 425 of this issue a group at the Universities of Turku and Jyväskylä, Finland, demonstrate that this ability is turned to deadly effect when hunting voles.

Voles communicate with one another by laying highways through grass and vegetation which they mark with urine and faeces. The world of a vole is dominated by smell: to a vole on its highway, a patch of scent-laden excrement is an informative as a road sign to a motorist. But as Viitala and co-workers show, kestrels can also read the signs and interpret them as measures of vole population density. It turns out that vole waste-products are strong absorbers of ultraviolet light, much stronger than the background of vegetation. Whereas to a human eye, the difference between a vole highway and any other patch of grass is not discernible, a kestrel sees every field as streaked with the marks of vole activity. In behavioural experiments both in the laboratory and in the wild, kestrels took a far greater interest in an environment daubed with ultraviolet-absorbing vole excrement than one without. This finding could explain how kestrels and other raptors always seem to be able to find their vole prey, even in the aftermath of the population 'crashes' so characteristic of northern rodents, when prey is generally scarce.

Henry Gee

J.P. Ferrero/Ardea

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

ing the time of basin formation must have varied strongly from one locale to another.

A new solution for global crustal thickness has been obtained² under the assumption that all of the variations in the measured gravity field are due to the known topography and to variations in the thickness of a low-density crust (see figure). Although this solution neglects the contribution to the gravity field of high-density mare basalt deposits, particularly those in the mascon maria, it does provide a first look at global patterns. The far side is characterized by thicker crust than the near side (68 versus 60 km), but by a lesser amount than in similar solutions obtained from earlier data lacking global coverage⁷. The range in crustal thickness is more than 100 km, and much of that range is to be found in the central far side.

More important is that the paucity of farside mare deposits is seen not to be simply the result of a greater elevation and thicker crust beneath the far side. Rather the small volumes of mare material in the deep depression of the South Pole–Aitken basin indicate that melt generation during the time of mare volcanism must have occurred much less readily in the mantle beneath the central far side than beneath the near side⁴. Together with the inference of an uneven nearside lithosphere thickness from the variable response to the emplacement of each mascon mare⁸, these results point to a strongly aspherical

temperature structure throughout much of the first 1,500 million years of lunar history. The challenge now is to unravel the respective contributions to this heterogeneity from the processes of lunar formation, later large impacts and mantle dynamics.

The first rush of data from Clementine has shown us a Moon that is much more complex in its thermal and mechanical evolution than had been appreciated. Additional findings are likely to be announced as more of the nearly two million digital images obtained during the mission¹ are sorted, calibrated and analysed. Doubtless the Moon has more remarkable secrets to reveal. □

Sean C. Solomon is in the Department of Terrestrial Magnetism, Carnegie Institution of Washington, Washington, DC 20015, USA.

1. Nozette, S. *et al.* *Science* **266**, 1835–1839 (1994).
2. Zuber, M. T., Smith, D. E., Lemoine, F. G. & Neumann, G. A. *Science* **266**, 1839–1843 (1994).
3. Spudis, P. D., Reisse, R. A. & Gillis, J. J. *Science* **266**, 1848–1851 (1994).
4. Lucey, P. G., Spudis, P. D., Zuber, M., Smith, D. & Malaret, E. *Science* **266**, 1855–1858 (1994).
5. Wilhelms, D. E. *U.S. Geol. Surv. Prof. Pap.* 1348 (1987).
6. Taylor, S. R. *Planetary Science: A Lunar Perspective* (Lunar Planet. Inst., Houston, 1982).
7. Kaula, W. M., Schubert, G., Lingenfelter, R. E., Sjogren, W. L. & Wollenhaupt, W. R. *Proc. 4th Lunar Sci. Conf.* 2811–2819 (Pergamon, New York, 1973).
8. Solomon, S. C. & Head, J. W. *Rev. Geophys. Space Phys.* **18**, 107–141 (1980).
9. Muller, P. M. & Sjogren, W. L. *Science* **161**, 680–684 (1968).

Which fault and what next?

Ross S. Stein

WHAT at first appears to be a semantic dispute between distinguished geologists about what to call Northridge's nemesis — the Pico or the Oak Ridge blind thrust fault — turns out to hinge on fundamental questions about where Los Angeles's most perilous fault might lurk^{1,2}. The only point of accord is that, by whatever name, the fault segment that slipped in last year's magnitude 6.9 earthquake is not a major player in southern California's pantheon of buried or 'blind' faults.

Writing in a November issue of *Nature*, Davis and Namson¹ argued that the fault on which the 17 January 1994 earthquake struck is a 'back thrust' or splinter off the more ominous Elysian Park blind thrust fault. The Elysian Park thrust is several times longer than the Pico, and in Davis and Namson's view, slips at three times the Pico's rate and is thus capable of much larger earthquakes. When the Elysian Park thrust ruptures, for example, the famed Hollywood sign will heave perhaps a metre heavenwards. Yeats and Huftile, writing on page 418 of this issue², argue

that the Northridge shock occurred on the east end of the Oak Ridge fault. They contend that the rate of slip of the Oak Ridge fault is three times that in the Ventura basin at Northridge and three times that of the Elysian Park thrust in Los Angeles. Thus they point to the Oak Ridge, which extends 70 km east from Northridge into the heavily populated Ventura basin and has shown a high rate of historical strain³, as the most ominous threat.

Davis and Namson's¹ key argument is that the Santa Susana anticlinorium north of the earthquake is a fold uplifted by slip on the blind south-dipping Pico thrust it conceals. The south flank of the fold rose by about half a metre during the earthquake⁴. Davis and Namson regard the north-dipping Santa Susana fault, whose trace snakes along the south flank of the anticlinorium, as a shallow bedding-plane thrust passively folded by the Pico fault. The fold, and thus the Pico below, is in their view about 30 km long, disappearing about

10 km east of the Oak Ridge fault, and is unrelated to it.

Yeats and Huftile² counter that a 60-mgal trough in the isostatic residual gravity field of the Ventura basin⁵ corresponds to the 15-km-deep sedimentary basin that abuts the downthrown side of the Oak Ridge fault⁶, and argue that since the gravity trough extends without interruption to the San Fernando Valley near Northridge, so must the Oak Ridge fault. So although the mapped fault trace disappears 30 km from Northridge, the fault continues by stealth as a blind thrust. Yeats and Huftile regard the Santa Susana anticlinorium as the product of slip on both the Oak Ridge and Santa Susana faults. In their view, the fold disappears west of the Oak Ridge trace because the Santa Susana fault, rather than the Pico, terminates at that point.

Both studies, to my mind, could go further to bolster their arguments. Davis and Namson do not show a map of the Santa Susana anticlinorium; its western terminus relative to the Oak Ridge and Santa Susana faults is the keystone to their contention that the Pico stops east of the Oak Ridge fault. Similarly, Yeats and Huftile do not show the gravity field. In fact the 60-mgal trough shallows to 48

EARTHQUAKES

From California to Kobe

THE Kobe earthquake offers a chilling vision of the damage that the Northridge shock might have wrought had it struck beneath Los Angeles. Despite the larger size of the Kobe event (moment magnitude $M_w=6.9$ rather than 6.7), the Northridge earthquake shook the ground harder. Two factors explain the heavier toll at Kobe. The Japanese shock ruptured towards and then through the port and industrial city. In contrast, the Northridge event struck beneath a suburban setting and ruptured away from Los Angeles. And the heavy industry and dense housing sited on soft soil and landfill at Kobe proved a tragic combination, whereas Northridge is underlain by stiffer soils. Unfortunately, Kobe's predicament is all too close to conditions elsewhere, such as along the Hayward fault that rims the east margin of the San Francisco Bay, where another of the world's great ports lies waiting. The Hayward fault produced two shocks of magnitude around 7 during the nineteenth century, and so, like Kobe, could rupture again.

Earth scientists in Japan have made impressive progress in analysing the earthquake, and have generously distributed preliminary findings through the Coordinating Committee for Earthquake Prediction, chaired by Kiyoo Mogi of Nihon University, to which we are indebted. The Kobe event struck on the

Arima-Takatsuki-Rokko fault system, one of the many shallow crustal faults in southwest Japan driven by the motion between the Philippine Sea and Eurasian plates. The fault lies 200 kilometres inland of the plate boundary, where two $M \approx 8$ earthquakes occurred between 1944 and 1946, and slips at about the same rate as the fault on which the Northridge earthquake struck, so earthquakes of magnitude 7 or over are rare. Seattle and Anchorage lie 250 and 350 kilometres, respectively, from their plate boundaries, and share some susceptibility to Kobe's fate.

The Kobe shock is a classic strike-slip earthquake, where each side of the fault moves sideways. The earthquake focus lies at the centre of an aftershock zone 50 kilometres long and 15 deep, a pattern reminiscent of the 1989 $M_w=6.9$ Loma Prieta earthquake on the San Andreas fault. As much as 2 metres of surface fault-slip was found on Awaji Island southwest of the epicentre, typical of strike-slip earthquakes of this size. Some 6–12 hours before the main shock, four foreshocks of magnitudes 1.5–3 occurred at the future epicentre, similar to the 1992 Landers, California, earthquake ($M_w=7.3$) and unlike the Loma Prieta shock. The Rokko strainmeter, a mere 25 kilometres northeast of the epicentre and within the aftershock zone,

however, shows no discernible precursor during the month or hours before the rupture, at a level of 0.1 microstrain. This is consistent with experience in the United States for shocks of similar size, but is nevertheless disappointing. Nor was any precursory magnetic signal seen at an instrument 50 kilometres from the epicentre.

In 1916 a shock of magnitude $M=6.1$ struck within a few kilometres of the 17 January 1995 epicentre. Small shocks cluster along the future rupture zone on Japanese maps of 1980–91 seismicity, reminiscent of many strikeslip faults in the US, such as the Hayward and the New Madrid seismic zone. The shear strain rate along the fault, also used in California to help gauge the earthquake hazard, is higher than average for Japan, but not by much. The rate, about 0.3–0.4 microstrain per year near the fault, is similar to the San Andreas and associated faults in the US. Fourteen years ago, in *Earthquake Prediction — An International Review* (Maurice Ewing Series 4, AGU, 1981), Tokihiko Matsuda placed the Arima-Takatsuki-Rokko fault among seven inland fault zones that appeared to be late in their earthquake cycles, and that he thus, with quite extraordinary foresight, termed 'precaution' faults.

R.S.S. and William L. Ellsworth, USGS

mgal and changes trend where the Oak Ridge disappears and the Santa Susana fold appears⁵, so the continuity of the trough is subject to interpretation. Each group would benefit from subjecting its cross-section to explicit gravity modelling.

More study of the fault segment that ruptured in 1994 is unlikely to resolve this dispute; the slip surface is already beautifully illuminated by its aftershocks⁴. Rather, what is needed is a closer look at the Santa Susana fault to the north and the Elysian Park fault to the south. The western termini of the Santa Susana fault and fold are crucial to arguments on both sides, and the slip rate of the Elysian Park thrust, which could range from 1.5 to 5 mm yr⁻¹, dictates whether a large earthquake on this fault — which straddles the heavily populated Los Angeles basin and the San Fernando Valley — could be in store. Ironically, calculations suggest that the Northridge earthquake increased the stress on both

the central Elysian Park and the eastern Oak Ridge faults⁷, and the rate of small shocks near both faults is now higher than it was before the Northridge earthquake struck. So, regardless of which fault is the more dangerous, both are showing signs of wear. □

Ross S. Stein is at the US Geological Survey, Menlo Park, California 94025, USA.

1. Davis, T. L. & Namson, J. S. *Nature* **372**, 167–169 (1994).
2. Yeats, R. S. & Huftile, G. J. *Nature* **373**, 418–420 (1995).
3. Donnellan, A., Hager, B. H. & King, R. W. *Nature* **366**, 333–336 (1993).
4. USGS & SCEC teams *Science* **266**, 389–397 (1994).
5. Roberts, C. W., Jachens, R. C. & Oliver, H. W. *Isostatic Residual Gravity Map of California and Offshore Southern California* (Calif. Div. Mines & Geol., Sacramento, 1990).
6. Jachens, R. C. & Griscorn, A. in *The Utility of Regional Gravity and Magnetic Anomaly Maps* (ed. Hinze, W. J.) 347–360 (Soc. Explor. Geophys., Houston, 1994).
7. Stein, R. S., King, G. C. P. & Lin, J. *Science* **265**, 1432–1435 (1994).

PALAEONTOLOGY

Between fish and amphibian

Robert Carroll

ONE of the most momentous events in the history of vertebrates was the emergence of terrestrial amphibians from their fish ancestors. Study of this transition has been frustrated by a gap in both time and morphology between the earliest known amphibians and their closest relatives among the fish. A new genus, *Elginerpeton*, described by Per Ahlberg on page 420 of this issue¹, provides new information that helps to fill that gap.

The fish most closely related to the amphibians are the osteolepiform sarcopterygians, which, like all terrestrial vertebrates, are characterized by having internal nostrils. Among the osteolepiforms, the genus that shares the most derived features with amphibians is *Panderichthys* from the lower part of the Upper Devonian (lower Frasnian, see figures). The best known of the early amphibians are *Acanthostega* and *Ichthyostega*, from the uppermost Devonian (upper Famennian), whose limbs appear to have been broadly similar to those of fully terrestrial vertebrates.

Skull and jaw bones of *Elginerpeton*, from rocks of upper Frasnian age in Scotland, exhibit a mosaic of fish and amphibian features. Some, such as an accessory tooth row on the dentary, are typical of osteolepiform fish, whereas others, including the pattern of ornamentation on the premaxilla and the lower jaw, and the presence of fang pairs and a tooth row on paired plates lateral to the jaw symphysis, are common to later amphibians. Bones of the upper limb

found in association with *Elginerpeton*² exhibit some tetrapod features, but whether this genus had feet like later amphibians, or fishlike fins, has not been established.

The most striking features of *Elginerpeton* and the contemporary genus *Obruchevichthys*³ from Latvia and Russia are unique derived characters that mean that they cannot be closely related to any of the previously recognized Upper Devonian or Carboniferous amphibians. The front of the skull is relatively narrow,

unlike those of both osteolepiform fish and uppermost Devonian amphibians, and the total cranial length is much greater. This family must represent a distinct, short-lived radiation, one which occurred before that which gave rise to most land vertebrates.

The discovery of *Elginerpeton* is but one of a number of finds during the 1990s that have contributed to understanding the fish/amphibian transition. Vorobyeva and Schultze, for example, have demonstrated⁴ that *Panderichthys*, although unquestionably a fish in its possession of paired fins, has many features that were thought to be restricted to amphibians. At the other end of the transition, Clack and Coates^{5–8} have shown that the amphibians *Acanthostega* and *Ichthyostega* retained many features indicative of an aquatic way of life. Two other late Devonian amphibians, *Hynerpeton*⁹ and *Tulerpeton*¹⁰, add to the overall picture in their possession of features of the post-cranial skeleton that are more advanced than the better known *Ichthyostega* and *Acanthostega*, suggesting closer affinities to land vertebrates of the Carboniferous and later.

Together, these fossils show that the transition between osteolepiform fish and land vertebrates did not occur within a single lineage, but encompassed a series of radiations beginning within a strictly aquatic environment, continuing through animals with an intermediate morphology, and culminating in lineages that are closely related to fully terrestrial animals. The diversity of intermediate forms reinforces Thomson's view¹¹ that this transition occurred within an extensive wetlands ecosystem that included a variety of distinct habitats.

Many of these fossils are incomplete, but enough of their skeletons is known to

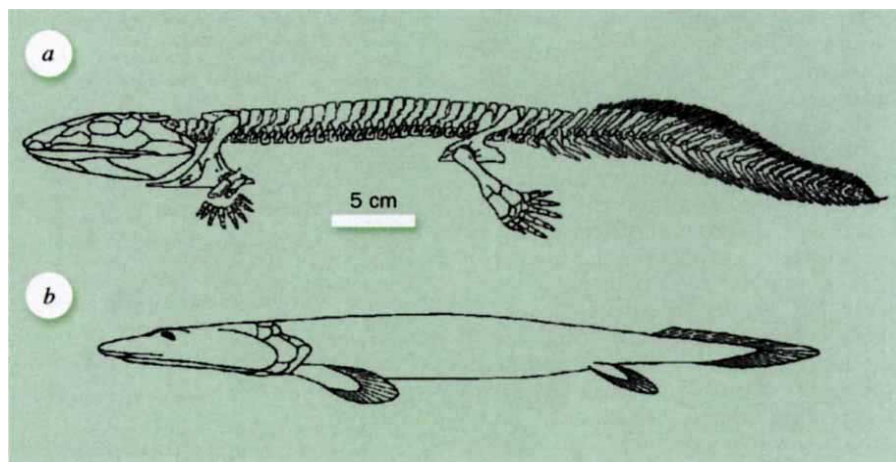


FIG. 1 *a*, Preliminary reconstruction of the late Devonian amphibian *Acanthostega* (redrawn from ref. 8). Note the presence of eight toes on the forelimb and seven on the rear. Later Palaeozoic tetrapods are not known to have more than five. The more proximal limb bones are basically similar to those of later terrestrial vertebrates. *b*, The lower Frasnian osteolepiform fish *Panderichthys* (redrawn from ref. 4). The paired fins may have functioned like those of other lobefinned fishes, but both the dorsal and anal fins have been lost, presumably in association with life in very shallow water.

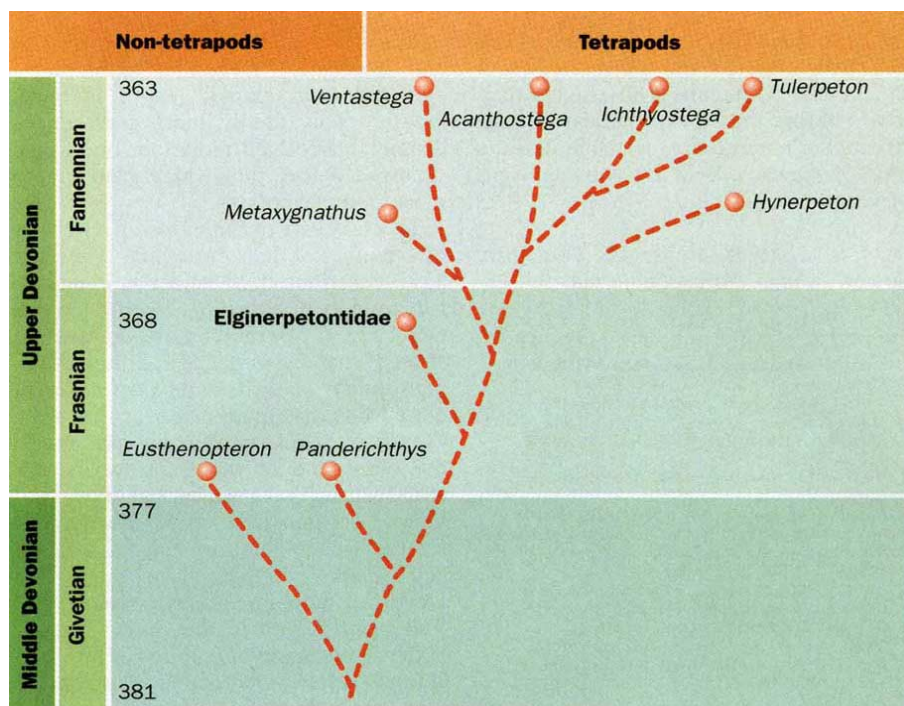


FIG. 2 The fish–amphibian transition. Temporal distribution of advanced osteolepiform fish and Devonian amphibians. Circles indicate time of occurrence of known fossils; dashed lines show possible relationships as discussed by Ahlberg¹; and the numbers indicate ages in millions of years.

indicate that skeletal features that were once thought to be unique to land vertebrates appeared stepwise in a succession of aquatic, semiaquatic and semiterrestrial vertebrates, while features commonly attributed to fish were retained in animals classified as amphibians. This point is highlighted by evidence, provided by Coates and Clack⁸, that important features of the tetrapod hands and feet in *Acanthostega* may have evolved among primarily aquatic animals. The distribution of primitive and derived characters differs from lineage to lineage, showing that many features were evolved or lost convergently. As in the case of other major transitions in vertebrates, such as the origin of birds and mammals^{12,13}, the convergent origin of derived features in different lineages makes it difficult to establish specific relationships, or to agree on objective criteria to differentiate tetrapods from their fish ancestors.

Although all these new data make it more difficult to classify species involved in the transition between fish and amphibians, they greatly illuminate the evolutionary processes involved. They demonstrate that the complete transition, from early osteolepiform sarcopterygians to primarily terrestrial amphibians may have extended over at least 15 million years. The fossil record is not yet well enough known to establish rates of evolution, but the observed changes may be attributed to progressive adaptation to increasingly shallow water, accompanied by repeated cladogenesis (branching of evolutionary lineages).

The increase of information on the fish–amphibian transition emphasizes the continued lack of knowledge of the next crucial step in the evolution of land vertebrates. In contrast to the variety of well-known Upper Devonian tetrapods, no amphibians have yet been described from the succeeding 20 million years, during which time all the main groups of later land vertebrates diverged. □

Robert Carroll is in the Department of Biology, McGill University, 1205 Dr Penfield Avenue, Montréal, Québec, Canada H3A 1B1.

1. Ahlberg, P. E. *Nature* **373**, 420–425 (1995).
2. Ahlberg, P. E. *Nature* **354**, 298–301 (1991).
3. Vorobyeva, E. I. *Trudy paleont. Inst.* **163**, 1–239 (1977).
4. Vorobyeva, E. I. & Schultze, H.-P. in *Origin of Higher Groups of Tetrapods: Controversy and Consensus* (eds Schultze, H.-P. & Trueb, L.) 68–109 (Cornell Univ. Press, Ithaca, 1991).
5. Clack, J. A. *Nature* **369**, 392–394 (1994).
6. Coates, M. I. & Clack, J. A. *Nature* **347**, 66–69 (1990).
7. Coates, M. I. & Clack, J. A. *Nature* **352**, 234–235 (1991).
8. Coates, M. I. & Clack, J. A. in *Proc. VIIth Symposium International Étude des Vertébrés Inférieurs* (in the press).
9. Daeschler, E. B., Shubin, N. H., Thomson, K. S. T. & Amaral, W. W. *Science* **265**, 639–642 (1994).
10. Lebedev, O. A. & Clack, J. A. *Palaeontology* **36**, 721–734 (1993).
11. Thomson, K. S. T. in *Major Features of Vertebrate Evolution* (ed. Spencer, R.) 85–107 (Paleontological Society, Dept Geol. Sci., Univ. Tennessee, Knoxville, 1994).
12. Ostrom, J. H. in *Major Features of Vertebrate Evolution* (ed. Spencer, R.) 160–177 (Paleontological Society, Dept Geol. Sci., Univ. Tennessee, Knoxville, 1994).
13. Hopson, J. A. in *Major Features of Vertebrate Evolution* (ed. Spencer, R.) 190–219 (Paleontological Society, Dept Geol. Sci., Univ. Tennessee, Knoxville, 1994).

DAEDALUS

Rolling foam

THE hovercraft supports itself on a cushion of air that leaks away so rapidly that it has to be replaced all the time by continuous fast pumping. Daedalus now proposes a hovercraft pumped not with air, but with detergent foam.

Foam, he reckons, acts like a gas that can sustain a pressure gradient. Every film wall in the foam has a pressure difference across it; the sharper the film curvature the greater this pressure drop. A foam of many small bubbles is a multitude of sharply curved films in series, and can contain quite a large pressure. Special detergent foams have already been proposed to impede rioters and reduce the force of terrorist bombs. With a little development, they should be able to support a hovercraft.

Daedalus's foamcraft will be essentially a flat platform with a hole in the middle. Into the hole will be pumped, not pure air, but detergent foam. Under the craft, it will spread out into a soft supporting cushion which will leak out round the edges only slowly. A very small pump will keep the vehicle foamborne.

The foamcraft could drive itself along in several ways. It could use the usual inefficient airscrews. A seagoing foamcraft could project a propeller shaft down through the foam into the water. On land it might use wheels or a caterpillar track, touching the ground but not loading it with much weight. While fitting the craft for its most obvious use — moving heavy loads over rough or marshy country — this would destroy much of its elegance. But Daedalus hopes that the foam itself can be made to exert a tractive force. By angling the jets delivering the foam, or spinning small paddle-wheels within it, it might be given a rotatory swirl which would turn it into a sort of fluid caterpillar track. The more viscous the foam, the greater traction it could exert, but at the cost of higher viscous losses. Careful optimization will be needed.

At first Daedalus imagined that his foamcraft would leave a glistening trail of foam behind it, like some giant mollusc. But he is now designing a simple foam-sucker at the back to scoop up this trail and recycle it. This would almost eliminate losses of water and detergent, making the craft nearly self-contained. Calmly and smoothly, with no blasts of air and no roaring, foamcraft will glide across water, along beaches, over marsh and moor and even along ordinary roads. Their cushion of churning detergent foam could make them ideal for street cleaning duties. Pedestrians run over by a foamcraft would not be injured, but merely subjected to a brisk Jacuzzi experience.

David Jones

The megalab truth test

SIR — Results from a new opinion poll suggest that lying in everyday life is more widespread than had previously been assumed¹. Only 12% of the sample claimed that they had never told a lie, whereas 24% claimed to have lied at least once during the previous day. It is thus worrying that a large body of research indicates that most individuals are unable reliably to detect when others are lying².

Much of this work has concentrated on identifying the cues used by observers to decide whether another individual is

of these interviews were printed in *The Daily Telegraph*, broadcast on BBC Radio 1 and shown on the BBC Tomorrow's World television programme. People were asked to decide which interview contained the lies and to telephone appropriate numbers to record their decision.

There was a huge response from the public ($n = 41,471$). Radio listeners detected the lies 73.4% of the time, newspaper readers 64.2%, and television viewers 51.8%. All three groups could detect the lies at above-chance levels (see table). Perhaps more interesting, there were significant differences between the detection rates of the three groups (d.f. = 2, $\chi^2 = 366$, $P < 1 \times 10^{-14}$). These differences support the notion that the presence of visual cues reduces individuals' ability to detect deception.

Clearly, the study was not perfect. The judgements were based on two relatively short interviews with just one person. The original design was to interview several politicians but this had to be abandoned due to the limited amount of broadcast time. Second, it was not possible randomly to allocate individuals to the three conditions. Third, one has to assume that the technology used to collect and collate a large number of responses in a short time was accurate and reliable. But despite these problems the study produced striking results and it is difficult to conclude that they are not due, at least to some extent, to the effect of the three media.

These results have important implications. Several studies have now shown that vocal and verbal cues tend to be more reliable indicators of deceit than visual cues¹. Despite this, observers watching videotapes of potential liars seem to base their decisions on visual cues. This may be because visual cues are more compelling than verbal/vocal ones or because observers falsely believe that such signals (for example, eye contact) are the best indicator of deceit. Whatever the explanation, it is clear that individuals wishing to detect deceit might be better off consciously paying more attention to verbal or vocal, as opposed to visual, cues.

The truth test was designed, in part, to increase public understanding of science. Although not perfect, the experiment illustrated how certain experimental methods (for example comparing performance between three conditions, 'blind' judging and so on) could be used to investigate an interesting and important issue.

In addition, the experiment received a large amount of media coverage³ and generated greater public involvement. Perhaps most important of all, it yielded interesting and useful data. That is assuming, of course, that the public were telling us the truth.

Richard Wiseman

Division of Psychology,
University of Hertfordshire,
College Lane,
Hatfield, AL10 9AB, UK

1. Wiseman, R. *The Daily Telegraph*, 24 (30 March 1994).
2. Miller, G. R. & Stiff, J. B. *Deceptive Communication*. (Sage, London, 1993).
3. Maier, N. R. & Thurber, J. *Personnel Psychol.* **21**, 23–30 (1968).
4. Bauchner, J.E., Brandt, D.R. & Miller, G.R. *Communication Ybk 1* (ed. Ruben, B. R.) 229–243 (Transaction, New Brunswick, New Jersey, 1977).
5. Highfield, R. *The Daily Telegraph* 26 (25 March 1994).

MEGALAB TRUTH TEST RESULTS			
	Newspaper	Radio	Television
No. correct	2,207	769	19,165
No. incorrect	1,231	279	17,820
% correct	64.2	73.4	51.8
χ^2 value (d.f. = 1)	277	229	49
P value	<1x10 ⁻¹⁴	<1x10 ⁻¹⁴	<1x10 ⁻¹⁴

All χ^2 and P values calculated by testing actual distributions against a theoretical 50/50 chance distribution.

telling the truth^{3,4}. Such cues fall into three broad categories: verbal cues, consisting of just the words used by the liar (including the number of words spoken, length of sentences); vocal cues, involving the way in which these words are said (voice pitch, pauses, hesitation); and visual cues, including any observable signs given off during the communication (eye contact, body movement, facial expressions). In a typical study, individuals are presented with films of liars and truth-tellers (containing all three types of cue), soundtracks of these films (containing vocal and verbal cues) or just the films' written transcripts (containing only verbal cues). The results have tended to be counterintuitive, with individuals watching the films (and therefore receiving the largest number of cues) exhibiting the lowest detection rates.

Many of these studies have been carried out with relatively small numbers of individuals (mostly university students) in laboratory environments. As a result, their conclusions may be applicable only to this rather limited set of circumstances. Despite the obvious importance of these results, no previous work had examined whether such findings generalize to a much wider cross-section of the population using actual newspapers, radio and television. Our 'megalab truth test' was designed to resolve this issue.

A well known British political commentator (Sir Robin Day) was interviewed twice about his favourite films. In one interview he consistently told the truth, in the other he consistently lied. Transcripts

Spores in earliest land plants

SIR — The earliest evidence for the colonization of the land surface by plants is in the form of desiccation-resistant spores¹. These spores predate taxonomically recognizable macrofossils by 30 million years. The material that makes up the walls (or exospores) of these spores — sporopollenin — maintains its fine structure over long spans of geological time. It is possible to document the ultrastructure of these dispersed spores using the electron microscope, and to infer relationships to modern plants by comparing their spore-wall ultrastructure.

I examined members of the dispersed spore genus *Dyadospora* (Lower Silurian Centerville Formation, Adams County, Ohio; Fig. 1). A cross-section of the exospore reveals a two-part construction (Fig. 2). An inner region, 1.4 μm thick when not deformed, is constructed of 10–15 distinct plate-shaped units or lamellae, each 50–150 nm thick. The outer region is 700–1,500 nm thick, and is not obviously lamellated.

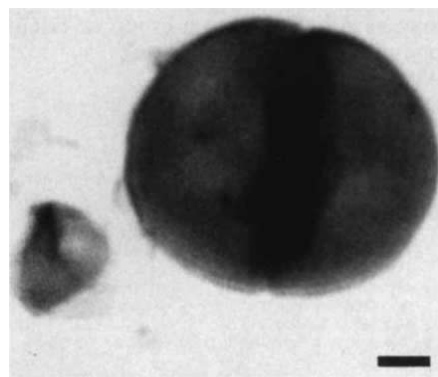


FIG. 1 *Dyadospora* sp. Light micrograph showing two adjacent spores of this dyad. Scale bar, 10 μm .

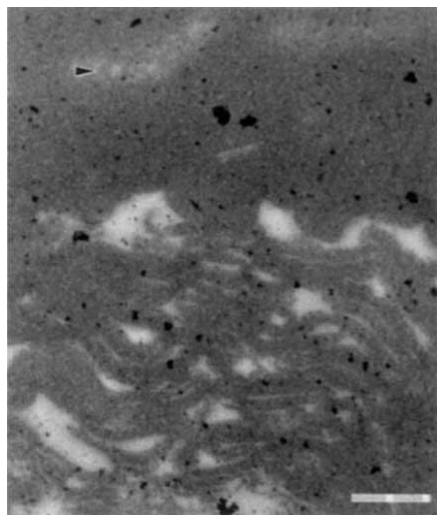


FIG. 2 Transmission electron micrograph showing a cross-section through the exospore of a single member of a dyad. The lightly stained area marked by the arrowhead is the spore surface. The material above this level is the outer region of the adjacent spore of this dyad. The material immediately below the arrowhead is the unlamellated outer exospore. The rest of the micrograph depicts the distinct lamellae of the inner exospore. Scale bar, 0.5 μm .

Exospores constructed of lamellae are nearly universal in modern land plant spores and pollen²⁻⁴. In most instances, the individual lamellae are highly fused at maturity, in some cases to the point of obscuring their separate status. In most, lamellated layers comprise a minority of the overall exospore thickness, but in hepatics, exospores consist almost exclusively of multiple, clearly separate lamellae. In particular, in the Sphaerocarpaceae there are several examples of multiple separate lamellae overlain by a more massive layer⁵. No other group of modern plants has this same combination of exospore ultrastructural features.

The lamellate structure of these spores suggests: (1) the early establishment of a developmental pattern common to all subsequent plant groups; and (2) the possibility that these early terrestrial pioneers may have belonged to a group of extant bryophytes.

Wilson A. Taylor

Department of Biology,
University of Wisconsin,
Eau Claire,
Wisconsin 54701,
USA

- Gray, J. Phil. Trans. R. Soc. B309, 167-195 (1985).
- Brown, R.C. & Lemmon, B.E. in *Microspores, Evolution and Ontogeny* (eds Blackmore, S. & Knox, R.B.) 55-94 (Academic, London, 1990).
- Rowley, J.R. & Southworth, D. *Nature* 213, 703-704 (1967).
- Blackmore, S. & Barnes, S.H. in *Microspores, Evolution and Ontogeny* (eds Blackmore, S. & Knox, R.B.) 173-192 (Academic, London, 1990).
- Steinkamp, M.P. thesis (Univ. California, Santa Cruz, 1973).

Larger bites of leaf-cutting ants

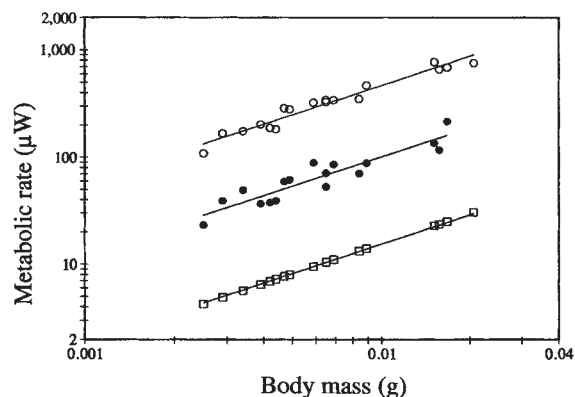
SIR — The maximum sustainable metabolic rate in non-flying animals ranges from about 8 to 12 times the resting or standard metabolic rate. This "aerobic scope" (maximum sustained metabolic/standard metabolic rate) increases to 20-100 for flying animals. Can scopes of this magnitude be attained without flight? To answer this question we investigated the leaf-cutting ant *Atta sexdens rubropilosa*, which cuts vegetation into convenient, portable fragments that it carries to the nest for processing by a resident fungal garden¹. By dissection and weighing we ascertained that the mandibular muscles in *Atta* comprise >50% of their head capsule mass, or >25% of total body mass — comparable to the flight motor of flying insects².

By impregnating Parafilm with citrus essences, we created a metabolically inert pseudo leaf that stimulated leaf-cutting behaviour in foraging ants³, whereupon we immediately placed the pseudo-leaf and attached, cutting worker in an ultra-sensitive flow-through respirometry system with high temporal resolution. Our sample of 18 ants (mean mass 7.89-5.41 mg, range 2.50-20.60 mg) cut at a mass-independent rate of 0.103-0.074 mm s⁻¹, equivalent to that reported for the cutting of tough or 'dense' leaves (0.1 mm s⁻¹, also mass-independent⁴). The leaf-cutting metabolic rate was stable and dramatically above both the standard rate and the post-cutting locomotion metabolic rate. We found no evidence for anaerobic metabolism during leaf-cutting; after leaf-cutting, the metabolic rate fell within 30 s to levels characteristic of post-cutting locomotion.

Both leaf-cutting and post-cutting metabolic rates were strong functions of body mass (see figure). The aerobic scope of post-cutting locomotion was 6.7 ± 1.4 , equivalent to the figure of 6.8 for unladen locomotion in *A. colombica*⁵. The aerobic scope of leaf-cutting (mean metabolic rate $49.2 \pm 6.8 \text{ W kg}^{-1}$) was far larger, at 30.7 ± 3.9 . Because of the near-identical mass scaling exponents of leaf-cutting metabolic rate and standard metabolic rates (see figure), this value is mass-independent.

By comparison, even the most prodigious feats of pedestrian locomotion yield

aerobic scopes of approximately 12 in vertebrates^{6,7} and invertebrates^{5,8}. Even flying insects, the most metabolically active animals known, attain aerobic scopes of only 20-100 (refs 2, 9; M. D. Dickinson and J. R. B. L., manuscript submitted). Leaf-cutting is obviously an energetically extraordinarily intense behaviour and constitutes a new record for metabolic activity not involving flight or flight-motor-associated endothermy. Indeed, when leaf-cutting metabolic rate is expressed specific to



Leaf-cutting metabolic rate (open circles) and post-cutting, locomotory metabolic rate (closed circles) as functions of body mass in *A. sexdens* at 25 °C. Both regressions shared a mass scaling exponent of 0.895; mass scaling coefficients were 28,741 and 6,180, respectively (overall $r^2 = 0.91$, $P < 0.0001$). The relation between body mass and estimated standard metabolic rate¹¹ is also shown (squares). Leaf-cutting metabolic rate (duration 1-7 min) was elevated 31-fold, and post-cutting rate 7-fold, above the standard rate. Flow-through, CO₂-based respirometry¹¹ used a Sable Systems TR-3 respirometry system. The respiratory quotient of attine ants during activity is characteristic of lipid catabolism⁴, corresponding to 27.6 J per ml CO₂, allowing us to calculate metabolic rate in SI energy units of μW or W kg^{-1} .

mandibular muscle mass it increases from 49.2 to $193.5 \pm 33.7 \text{ W kg}^{-1}$, a figure approaching that for insect flight muscle, the most metabolically intense tissue known ($300-3,000 \text{ W kg}^{-1}$).

Thus, mandibular energetics may play an unexpectedly important role in ant foraging. Specifically, the effects of mandibular energetics on load-size selection and foraging efficiency (at the individual level) and on the occurrence and intensity of recruitment (at the colony level) have probably affected the evolu-

- Hölldobler, B. & Wilson, E.O. *The Ants* Harvard Univ. Press, Cambridge, Massachusetts, 1990).
- Casey, T.M. & Ellington, C.P. In *Energy Transformations in Cells and Organisms*, (eds Wieser, W. & Gnaiger, E.) 200-210 (Thieme, Stuttgart, 1989).
- Roces, F. *Anim. Behav.* 40, 1181-1183 (1990).
- Roces, F. & Hölldobler, B. *Oecologia* 97, 1-8 (1994).
- Lighton, J.R.B., Bartholomew, G.A. & Feener, D.H. *Physiol. Zool.* 60, 524-537 (1987).
- Pasquis, P.A., Lacaille, A. & Dejourn, P. *Respir. Physiol.* 9, 298-309 (1970).
- Taylor, C.R. et al. *Respir. Physiol.* 44, 25-37 (1981).
- Full, R. J. in *Energy Transformations in Cells and Organisms* (eds Wieser, W. & Gnaiger, E.) 175-182 (Thieme, Stuttgart, 1989).
- Ellington, C.P. *J. exp. Biol.* 160, 71-91 (1991).
- Wilson, E.O. *Behav. Ecol. Sociobiol.* 7, 157-165 (1980).
- Lighton, J.R.B. & Wehner, R. *J. comp. Physiol.* 163, 11-17 (1993).

tion of worker polymorphism and foraging strategies in leaf-cutting ants^{1,10}, and may be significant in other ant genera.

Flavio Roces

Theodor-Boveri Institut,
Lehrstuhl für Verhaltensphysiologie
und Soziobiologie der Universität,
97074 Würzburg, Germany

John R. B. Lighton

Department of Biology,
University of Utah, Salt Lake City,
Utah 84112, USA

Nuclear DNA from primate dung

SIR — Although hypervariable DNA markers have revolutionized the study of kinship patterns and mating systems in the wild^{1,2}, these techniques have largely been restricted to species that can be easily captured and handled. Where handling is either unsafe or impracticable, DNA must be obtained non-invasively³. Collecting hair or buccal cells can be difficult or impossible in many species, so the ideal system should be based on material that is both plentiful and can be attributed unambiguously to a specific individual. Following the extraction of mitochondrial and chloroplast DNA from bear dung, by Höss *et al.*⁴, we investigated whether useful samples of nuclear DNA could be obtained from the faeces of wild primates.

We collected dung from individually recognized olive baboons in the Gombe

Stream National Park, Tanzania^{5,6}, and stored it first in liquid nitrogen, then in a -20°C freezer. We modified a protocol for extracting mitochondrial DNA from chimpanzee dung⁷, combining the techniques of Pääbo⁸ and Boom⁹ with a phenol/chloroform extraction with hexadecyltrimethylammonium bromide (CTAB) and collagenase. CTAB was added to the overnight incubation in order to break down ingested plant compounds that would inhibit subsequent PCR reactions. Extracted DNA was amplified using human nuclear microsatellite primer D4S243 (ref. 10) and PCR. As a control, we used chelex³ to extract DNA from hair of several of the same individuals. PCR products were inserted into TA cloning vector (Invitrogen) and transformed into *Escherichia coli*. Colonies were grown overnight, plasmids were isolated and sequenced with the M13 forward primer.

Nuclear DNA extracted from dung and hair reveals that the baboon sequence at the D4S243 locus aligns with the human sequence, exhibiting three base substitutions, four ambiguous codes, and deletions of 28 and 63 base pairs (see figure). Human microsatellite primers successfully amplify baboon DNA in a significant number of cases (J. Rogers, personal communication), and paternity exclusions can usually be achieved by amplifying six polymorphic loci. PCR amplification has been successful in 39 out of 40 faecal samples tested so far. The ability to amplify microsatellite regions from faecally extracted nuclear DNA

opens a new realm of possibilities for studying organisms that are difficult to catch or handle¹¹. Dung should be considered essential material for paternity and population genetic studies of arboreal or endangered mammals.

J. J. Constable*, C. Packer,

D. A. Collins, A. E. Pusey

Department of Ecology,
Evolution and Behavior,
University of Minnesota,
St Paul, Minnesota 55108, USA

*Present address: Department of Biology, University of Nevada, Mailstop 314, Reno, NV 89557, USA.

1. Burke, T. *Trends ecol. Evol.* **4**, 139 (1989).
2. Queller, D.C., Strassman, J.E. & Hughes, C. R. *Trends ecol. Evol.* **8**, 285 (1993).
3. Morin, P.A. *et al. Science* **265**, 1193 (1994).
4. Höss, M. *et al. Nature* **359**, 199 (1992).
5. Goodall, J. *The Chimpanzees of Gombe* (Harvard Univ. Press, Cambridge, 1986).
6. Packer, C. *et al. Nature* **373**, 60–63 (1995).
7. Goldberg, T. *Biotechniques* (in the press).
8. Pääbo, S., Gifford, J.A. & Wilson, A.C. *Nucleic Acids Res.* **16**, 9775–9787 (1988).
9. Boom, R. *et al. J. clin. Microbiol.* **28**, 495–503 (1990).
10. Hudson, *et al. Genomics* **13**, 622 (1992).
11. Kohn, M. *et al. Molec. Ecol.* (in the press).

Comparison of deep ice cores

SIR — Comparison of the GRIP and GISP2 deep ice cores from central Greenland^{1,2} has confirmed the occurrence of exceptionally large, rapid changes in many climatic indicators over approximately the past 100,000 years^{3,4}. Similar rapid changes occur within deeper ice with a warm isotope signature (identified as the Eemian, Sangamonian, or stage 5e^{3,4}), but differences in their patterns between the two cores raise the possibility that at least one record was disturbed by ice-flow processes^{1,2}. Disturbances might include boudinage, open folding or similar processes that would distort the time-depth relation but leave layers in stratigraphic order, or overturned folding that would disturb stratigraphic order. The GRIP and GISP2 steering bodies have initiated comparative studies of the two cores, including visible stratigraphy and crystal fabrics in selected sections of both cores, which are reported here. The most important results of these studies include the following.

(1) The small diameters of the cores and the lack of any highly reliable stratigraphic 'up' indicators prevent us from distinguishing overturned from right-side-up layers in structures much larger than the core diameter (13 cm for GISP2 and 10 cm for GRIP before sampling). We thus cannot invalidate or confirm the instability of the Eemian record.

(2) Both cores contain structures above the Eemian (which was identified as 2,790–2,865 m at GRIP^{3,4}) that are larger than the core diameter and that could represent inverted strata. A 10–20-cm long region of layers dipping 20° relative

a
CTAG

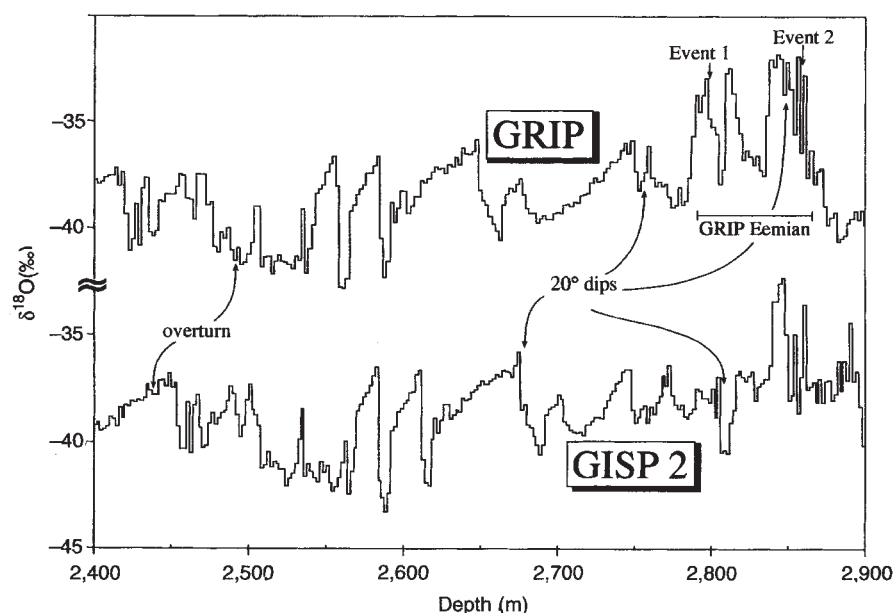


b

Human:	TAGGAGCCTG	TGGTCTCTGT	GGTGTAATT	GTATTAGGAA	GAGAGGAGAG
Baboon:	N.....	N.....	N.....	G.....	N.....
	ATAAAGATG	TAAATGGAGA	TCIGTCTGTC	TACCTATCTA	TCTATCCGTT
	G.....	G.....	G.....	G.....	G.....
	TATCTATCTA	-----	-----	TCTATCTATC	TATCTATCTA
	-----	TCTATCTATC	TATCTATCTA	-----	-----
	TCTATCTATC	TATCATCTAT	CTGCAAGGAG	AAAGAGAGAC	TGAAGAGAAT
					
	GCAGGGTGAT	AGAAAGGTAG	AAAAGGGATT	GAAATCATTG	TTAAAAGGAA
	-----	-----	-----	-----	-----
	TGAGAA	-----	-----	-----	-----

a, The D4S243 locus of a baboon, sequenced from the PCR amplification of faecal DNA. The fragment length is 169 base pairs, with a prominent tetra-repeat of 16 times. PCR amplification utilized a 40-cycle "touch-down" program, terminating at 50°C. Sequencing was achieved by inserting the PCR product into Invitrogen's TA cloning vector, and reactions were performed using Promega's cycle sequencing kit and a 6% denaturing acrylamide gel for 3 h at 55 W. Visualization was accomplished with Promega's silver stain kit.

b, The human D4S243 locus aligned with the extracted baboon sequence. Note the 28-base-pair deletion in the baboon sequence before the repeat region and the 63-base-pair deletion at the end of the fragment. The length of the repeat region may vary between individuals, thus allowing paternity and population-genetic studies.



Oxygen-isotope compositions of the GRIP⁴ and GISP2 (ref. 2) deep ice cores. The shallowest overturned folds, and the shallowest two regions of steep dips, are indicated for each core. The depths of major features in the cores differ slightly owing to well-understood differences in snow accumulation and layer thinning; significant mismatch not explainable in this way occurs deeper than about 2,700–2,750 m.

to the normal to the core axis, within a zone more than 1 m long of dips of 10° or more (compared with a borehole inclination of about 2°), occurs at about 2,757 m in the GRIP core, and steep dips also are observed near 2,847 m. In the GISP2 core, a 50-cm-long region with dips as steep as 20° was observed at 2,679 m, and a region several metres long with dips to more than 20° occurs around 2,809 m, with other regions of steep dips deeper (compared with a borehole inclination of about 5°). We cannot yet guarantee that these are the shallowest occurrences of disturbed features too large to be observed completely within a single core. Most of the ice around these features in both cores is less disturbed, although much disturbance is evident below the GRIP Eemian and below a similar level at GISP2. Significant mismatch between the palaeoclimatic records begins at about the same depth as the large deformational structures, but this correspondence is not exact and some deformation occurs without large mismatch.

(3) Both cores contain evidence of at least moderate simple-shear deformation (strains of the order of 10 or more), including z-shaped folds with overturned limbs smaller than the core diameter. Such z-shaped folds, millimetres high, appear as shallowly as 2,437 m at GISP2 and 2,483 m at GRIP with larger folds deeper, but those at GRIP do not appear as well-developed as those at GISP2 until the lower quarter of the Eemian (2,847 m and below, including the margins of event 2; refs 2,3). The shear sense is consistent within each core. In both cores, the c-axis fabrics are affected by folding and regions

of dips, indicating that deformation is ongoing or sufficiently recent (within about 3,000 years) that the fabric has not been reset by further deformation. GRIP today is near the ice divide, where flow patterns are not expected to produce stratigraphic discontinuities, whereas GISP2 is in a simple-shear, flank-flow regime where stratigraphic discontinuities are more likely. The simple-shear features observed at GRIP suggest recent or ongoing ice-divide migration^{5,6}.

R. B. Alley*

A. J. Gow†

S. J. Johnsen‡

J. Kipfstuhl§

D. A. Meese†

Th. Thorsteinsson¶

*Earth System Science Center
and Department of Geosciences,
The Pennsylvania State University,
University Park, Pennsylvania 16802, USA

†US Army Cold Regions,
Research and Engineering Laboratory,
Hanover, New Hampshire 03755, USA

‡The Niels Bohr Institute,
University of Copenhagen,
DK-2200 Copenhagen N, Denmark

§Alfred-Wegener Institut für
Polar und Meeresforschung,
Columbusstrasse,
27568 Bremerhaven, Germany

¶To whom correspondence should be addressed.

1. Taylor, K. *et al.* *Nature* **366**, 549–552 (1993).
2. Grootes, P. *et al.* *Nature* **366**, 552–553 (1993).
3. GRIP Project Members. *Nature* **364**, 203–207 (1993).
4. Dansgaard, W. *et al.* *Nature* **364**, 218–220 (1993).
5. Hempel, L. & Thyssen, F. *Polarforschung* **62**, 6–11 (1992).
6. Anandakrishnan, S., Alley, R.B. & Waddington, E.D. *Geophys. Res. Lett.* **21**, 441–444 (1993).

Back to front

SIR — Lim *et al.*¹ and Feng *et al.*² showed that Src-homology-3 (SH3) domains can bind two classes of ligand that exist in opposite orientations on the protein-carbon backbone. These elegant studies represent an example of a protein-protein interaction in which the ligand sequence can exist in either N→C or C→N orientation. We have also demonstrated that antibody molecules can similarly recognize protein sequences presented in either orientation³.

These various findings should sound a general warning to the scientific community regarding the use of protein databases for similarity searches to detect related binding motifs. Clearly, such searches should be performed in both forward and reverse carbon-backbone orientations; otherwise, up to half of the functional homology regions could be overlooked.

Dennis W. Metzger

Department of Microbiology,
Medical College of Ohio,
Toledo, Ohio 43699-0008, USA

Victor H. Van Cleave

Genetics Institute, Cambridge,
Massachusetts 02140, USA

1. Lim, W. A., Richards, F. M. & Fox, R. O. *Nature* **372**, 375–379 (1994).
2. Feng, S., Chen, J. K., Yu, H., Simon, J. A. & Schreiber, S. L. *Science* **266**, 1241–1247 (1994).
3. Van Cleave, V. H., Naevie, C. W. & Metzger, D. W. *J. exp. Med.* **167**, 1841–1848 (1988).

Chromosome correction

SIR — Reese *et al.* (*Nature* **371**, 523–527; 1994) published the isolation of two yeast TBP-associated proteins. One of them, named by the authors yTAFII90, corresponds to a gene that we have previously sequenced (L. M. *et al.* *Yeast* **10**, 819–831; 1994). In our paper, we mentioned the strong similarity of YBR1410 with the TAFII80 from *Drosophila* which was at that time the only known sequenced gene of this family. Reese *et al.* used our sequence present in the database, but they erroneously referred to chromosome III sequencing. Thus, we want to correct this point: YBR1410/yTAFII90 belongs to chromosome II (accession number S34023).

Françoise Bussereau

Nadine Démolis

Laurent Mallet

Michel Jacquet

Institut de Génétique et Microbiologie,
CNRS URA 1354,
Université de Paris-Sud,
Bâtiment 400,
91405 Orsay Cedex,
France

Symptoms or signs?

Roy Porter

Approaching Hysteria: Disease and its Interpretations. By Mark Micale. Princeton University Press: 1995. Pp. 327. \$29.95, £24.95.

HYSTERIA, mysteria: thus ran a popular quip about the bizarreness of the affliction that laid low so many Victorian women with its strange repertoire of symptoms, from fainting fits and the megrims up to apparent spinal paralysis, and which itself seems to have disappeared with mystifying suddenness early in the present century. Patients presenting with the florid symptoms that were bread-and-butter to Professor Charcot, that "Napoleon of the Neuroses", in *belle époque* Paris or to the young Sigmund Freud in Vienna are now, it seems, extinct, at least in the Western world; hysteria has been dropped from most modern diagnostic manuals; and surely for related reasons, hysteria long ago ceased to be an object of scholarly interest.

But that has changed: right now it is the flavour of the month. Studies pour off the presses of virtuoso sufferers (such as Anna O., treated by Freud and Breuer) and of celebrated nerve doctors, including Pierre Briquet, whose work is now considered so important that there are moves afoot to rename hysteria 'Briquet's syndrome'. The reasons are clear. Freud remains as provocative and perplexing as ever. New and apparently inexplicable disorders grow worse (anorexia nervosa, myalgic encephalomyelitis), paining patients, puzzling physicians and driving us back for illumination to the enigmatic maladies of the past. And, above all, over the past couple of decades, feminists have shed new light on the ways in which hysteria formed both the battleground and the weaponry for women's struggles in patriarchal societies. Even so, the quandaries remain: should *la grande hystérie* be seen primarily as a symptom of women's doll's-house oppression, as a scream of protest or as an (unconscious) act of political defiance?

Readers seeking a lucid and scrupulously balanced introduction to current debates on this puzzling condition, both in the medico-psychiatric community and particularly among historians, sociologists and literary critics, will be deeply grateful to Mark Micale. His *Approaching Hysteria* affords a comprehensive mapping of recent avenues of inquiry, evaluating the work of hundreds of scholars who have contributed over the past 20 years to the revival of inquiry into hysteria. He has mastered the medical doctrines of yesteryear, while, as a professional historian, adroitly setting them, together with evidence from autobiographies and fiction,

into their wider intellectual contexts. Although often critical, Micale avoids grinding partisan axes or riding pet hobby-horses. He is intent, however, to dispel certain hoary myths. It is wrong to suppose, for instance, that hysterics were exclusively rich and pampered; recent studies have shown that many of the women treated for hysteria in the eighteenth and nineteenth centuries were servants, waitresses and other working women: did they, one wonders, learn hysterical behaviour from their betters?

tribute. He demonstrates that, especially among nineteenth-century Parisian physicians, the reality of male hysteria was well established. The male and female varieties clearly remained demarcated by gender, however, as was made especially plain in the First World War once shell-shock became accepted as a form of hysteria.

Micale also shows how the term hysteria became, particularly at the *fin de siècle*, a kind of universal, catch-all cultural metaphor. Often used to evoke a sense of decadence and cultural degeneration, it served above all as a shorthand for woman herself. "Hysteria, madam, now there is the word of the day", remarked Guy de Maupassant. "Are you in love? You are a hysteric. Are you indifferent to the passions that arouse other people? You are a hysteric, but a chaste hysteric. Do you cheat on your husband? You're a hysteric, but a sensual hysteric. Do you steal

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Mass hysteria — 'hysterical' patients simultaneously put into hypnotic trances by a loud unexpected sound. From Regward's *Maladies de l'esprit* (1887).

It is one of Micale's main contentions that hysteria was too intricate a condition to allow for any simple or reductionist interpretation. There was obviously a psychological component to it, but that does not mean that some Victorian 'hysterics' wouldn't now correctly be given organic diagnoses. Similarly, it is true that hysteria was mainly a disease-label pinned on women by male physicians. But it does not follow that the diagnosis was just some crass patriarchal ruse, or that certain females didn't seek, and obtain, some 'secondary gain' by (consciously or unconsciously) adopting the role of the hysteric. Hysteria thus defies easy historical answers, and Micale leads us admirably through the thickets, with just an occasional rap over the knuckles for those (mainly postmodernist feminist literary critics) who think they have discovered the key to the mystery.

As well as surveying the work of others, Micale has insights of his own to con-

tribute. He demonstrates that, especially among nineteenth-century Parisian physicians, the reality of male hysteria was well established. The male and female varieties clearly remained demarcated by gender, however, as was made especially plain in the First World War once shell-shock became accepted as a form of hysteria.

But it does not follow that the diagnosis was just some crass patriarchal ruse, or that certain females didn't seek, and obtain, some 'secondary gain' by (consciously or unconsciously) adopting the role of the hysteric. Hysteria thus defies easy historical answers, and Micale leads us admirably through the thickets, with just an occasional rap over the knuckles for those (mainly postmodernist feminist literary critics) who think they have discovered the key to the mystery.

Roy Porter is in the Wellcome Institute for the History of Medicine, 183 Euston Road, London NW1 2BN, UK.

Towards a new physics of mind

Colin McGinn

Mental Reality. By Galen Strawson. MIT Press: 1994. Pp. 337. \$34.95, £24.95.

POSITIVISM may be officially dead but it enjoys a thriving afterlife in much contemporary psychology and philosophy of mind. Its chief manifestation is a persistent tendency towards behaviouristic views of the mind: mental phenomena are conceived as essentially involving motions of the body or dispositions to such motions. The root reason for this doctrine is the metaphysical principle that all reality is publicly observable reality. Thus pain, say, which is commonsensically taken to be a modality of private inner experience defined solely by its phenomenal character, is edged in the direction of the public domain by reducing it to writhings and groanings and reportings. What are normally regarded as the effects of pain are introduced into its essential nature. In this way pain becomes open to perceptual scrutiny, and hence scientific treatment.

Galen Strawson attacks this legacy of positivism, offering an entirely nonbehaviouristic view of the mind. His most vivid technique is to describe an alien species of "Weather Watchers" who are immobile and entirely without volitional capacity. They wait and watch, see and hear, hope and fear — but they never do anything. Nor are they disposed to do anything, except in the trivial sense that anything is disposed to do anything, if one is allowed to change the nature of the thing in question enough. As Strawson tellingly remarks, a table is disposed to be a mathematician, if we are allowed to say that it would do mathematics if its parts were rearranged so as to copy a human brain. The point about the Weather Watchers is that they have a recognizable mental life although they lack the ability to act. And if this is so — if, that is, such a race of beings is logically possible — then it cannot be part of the very concept of a mental state that contains reference to behaviour. All we have really done in imagining this case is to subtract one sort of contingent effect from mental states as we ourselves exemplify them; we have not removed the essence of the state. But if this is right, then vast tracts of contemporary thought about the mind — from Wittgensteinians to functionalists — are simply wrong about the nature of mental phenomena. The ghost of positivism must be finally laid to rest: it has been misbehaving for too long.

But what are mental phenomena if they are not behaviourally constituted? Strawson answers that they are physical phenomena, occurrences in the brain. This

might seem an odd view, given the marked difference between conscious experiences and run-of-the-mill neural states. Couldn't we also imagine a race of beings who have neural states but no experiences? But this is to misunderstand Strawson's thesis, which is that experiences are a new sort of physical state, not recognized in current or foreseeable physics. Experience is real; materialism is true; nothing in current physics covers experiential properties of the world; therefore there must be physical properties of which physics is so far ignorant. Strawson calls this position "agnostic materialism": we know that materialism is true, but we don't have any idea of what kind of physical theory might capture the nature of mental properties. Accordingly, today's physics is radically incomplete, because it fails to include any theory of one of the most salient and certain characteristics of physical reality, namely conscious experience. A genuinely unified theory of physical reality would have to find a place for consciousness, linking it to other aspects of the physical world; but such a theory is not even a glimmer in the eye of the physicist. At present, indeed, physics simply ignores consciousness, thereby dodging a problem it should try to find the resources to explain. (Roger Penrose is a physicist who takes a similar view of the problems posed for contemporary science by consciousness.)

Strawson supports his materialist metaphysics with a subtle and unusual assessment of the traditional rivals to materialism: idealism and immaterialism. Materialistic scientists may find some of this abstract conceptual investigation not to their taste, although it is important to see that materialism is just as much of a metaphysical position as those less prevalent systems of thought. Positivism was also wrong in supposing that metaphysics is an avoidable (and meaningless) philosophical indulgence: some general view of reality must be assumed even in the conduct of hard science. In any case, there are, it turns out, good arguments that favour materialism over its traditional rivals, so it is not merely a dogma into which we have uncritically slipped.

Strawson also kicks against the pricks in his discussion of intentionality — the 'aboutness' of mental states. His thesis is that there is no deep problem of intentionality over and above the earlier problem of experience. Here I think he underestimates the problematic character of intentionality. First, he fails even to mention the main difficulty with causal-naturalistic theories: that they cannot account for the possibility of *mis*-representation. Second, and more importantly, he seems to suppose that all intentionality is extrinsic to the qualitative character of experience, being superadded by means of causal-type relations to the world. But it is

more plausible to take experiential content to be *constituted* by intentionality, so that it is integral to solving the experience problem. How else are we to characterize an experience of red except as an experience of red? Nevertheless, this is a challenging discussion that deserves to be taken seriously.

The view of mind with which Strawson ends up is described as commonsensical. It holds that experiences are real, ontologically independent of behaviour, unobservable, adequately conceived by folk psychology and possessed of genuine intentionality. It also holds that they are the best candidates for the epithet 'mental'. He is clearly right that this position is part of common sense, despite the very widespread rejection of it in twentieth-century psychology and philosophy. He is also right, I think, in saying that no cogent reasons have ever been given for rejecting common sense on this matter. If so, then we are at last at the point where serious study of the mind can begin: we have at least correctly identified our subject matter. The question is whether we shall be able to forge any satisfying theories for mental phenomena once they are rightly conceived. How *does* an experience of red find its way into a world of atoms and molecules and cells? That old question stands out in all its naked profundity.

Strawson's book is a careful, sensitive and imaginative treatment of some of the main conceptual questions that condition any approach to the nature of mind. If it is sometimes a little long-winded and repetitive, and occasionally belabours the obvious, then we need to remember that it flies in the face of a tremendous amount of received philosophical and scientific doctrine. *Mental Reality* has the rare and notable virtue of being not fundamentally wrong. □

Colin McGinn is in the Department of Philosophy, Rutgers University, Davison Hall, Douglass Campus, New Brunswick, New Jersey 08903, USA.

New in paperback

In the Wake of Chaos: Unpredictable Order in Dynamical Systems by Stephen H. Kellert. University of Chicago Press, \$12.50, £8.75. A philosophical dissection of chaos theory and its implications for science as a source of knowledge and how we see the world in general.

The Prehistoric Exploration and Colonisation of the Pacific by Geoffrey Irwin. Cambridge University Press, £13.95, \$17.95. The author uses, among other things, computer modelling to argue that Polynesian voyaging was rapid, purposeful and systematic and that navigation methods progressively improved.

Dividing the mind

Michael C. Corballis

Brain Asymmetry. Edited by Richard J. Davidson and Kenneth Hugdahl. MIT Press: 1995. Pp. 735. \$75, £67.50.

WHY is cerebral asymmetry so relentlessly fascinating? In one of the chapters in this volume, Albert M. Galaburda hints at a possible answer: "Despite the fact that the left hemisphere is significantly different in function from the right, there appears to be no structure or chemical constituent that is present in one hemisphere but not in the other." This discrepancy between function and structure is reminiscent of Descartes' notion of a nonmaterial mind that transcends the physical brain, and, because the functional asymmetry is especially pronounced in humans, it may even suggest a possible basis for human uniqueness.

The left-cerebral dominance for language in humans was discovered in the nineteenth century, and prompted much speculation as to its deeper significance. This speculation, amusingly documented by Anne Harrington in the opening chapter, became increasingly exaggerated, verging even on the occult, and interest lapsed early in the twentieth century. It was revived in the 1960s following the striking demonstrations of cerebral asymmetry in split-brained patients. Harrington's chapter stands as a timely warning, as over-fanciful speculation again threatens to drive the topic into oblivion.

This collection, however, represents the solid scientific front of the laterality enterprise, although there is perhaps a hint of desperation in the search for coherence in the now bewildering and often conflicting array of findings. Halle D. Brown and Stephen M. Kosslyn, for example, reviewing evidence on asymmetries in visual processing, write that "The difficulty in obtaining laterality effects at all, and particularly the failure to replicate findings across studies, should be the focus of future research." The sceptical reader may wonder. One generalization that has stood the test of time reasonably well, though, is that the left brain tends to operate in an analytical way, in contrast with the more holistic style of the right. The chapters by Brown and Kosslyn and by Joseph B. Hellige examine variants of this idea, with some interesting speculation about its neural and developmental bases.

It has also become clear that cerebral asymmetry is not after all exclusive to humans. Galaburda points out that we now know more about structural asymmetries in the nonhuman than in the human brain, although functional dominance remains more strikingly evident in humans. David Warren Lewis and Marian

Cleeves Diamond document evidence for hormone-dependent sexually dimorphic cortical asymmetries in rats, reminiscent of the late Norman Geschwind's suggestion that hormone levels also govern the development of cerebral asymmetry in humans. The link remains speculative, but seems certain to remain a focus of inter-species research for some time.

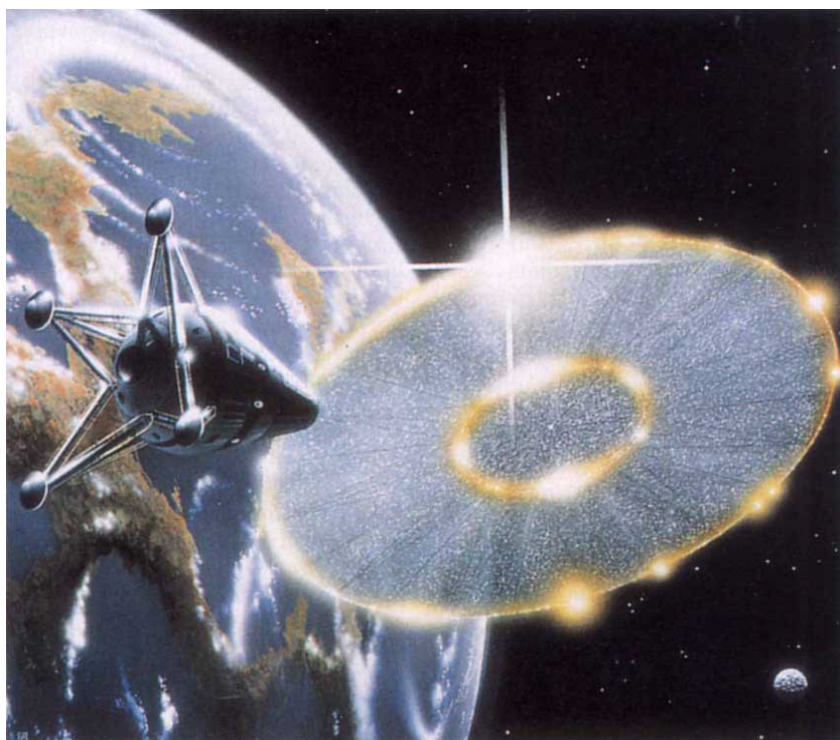
Merrill Hiscock and Marcel Kinsbourne provide a crisp review of the phylogeny and ontogeny of cerebral asymmetry, but resist evolutionary accounts that stress the adaptive nature of asymmetrical representation on the grounds that the bilateral representation of language does not impair its efficacy. They overlook the idea, first attributed to the American physician S. T. Orton in the 1920s, that lack of cerebral dominance may underlie language disorders such as dyslexia and stuttering. This has proved to be controversial, but chapters by Galaburda, George W. Hynd and colleagues, and Carol A. Boliek and John E. Obrzut amply document the reversal or lack of asymmetry in dyslexics. Reading itself cannot have been decisive in the evolution of cerebral asymmetry, but it may nevertheless reflect deeper language skills that were.

Other chapters suggest that anomalies of cerebral asymmetry may underlie pathologies such as mania, anxiety disorders, psychosis (Gerard E. Bruder; Robert G. Robinson and Jack E. Downhill) and

even sudden cardiac death (Richard D. Lane and J. Richard Jennings). These attributions may reflect in part an age-old tendency to bestow inferiority on those who deviate from the norm; at one time, the primary victims were left-handers, but with the discovery of cerebral asymmetry the net has been cast more widely. Even so, the evidence is sufficiently pervasive to suggest that speculation about the adaptive nature of cerebral asymmetry may not be out of place.

The most frequently cited author in the book is Justine Sergent, who died tragically last April. Her legacy in the understanding of visual asymmetries is acknowledged in the chapters by Brown and Kosslyn and by Hellige, and her work on cerebral asymmetries in the processing of faces is summarized in her own chapter. Her research on interhemispheric integration in the split brain forms the basis of attempts by Marie T. Banich and by Jacqueline Liederman to reinterpret the split-brain syndrome. These revisions are in my opinion premature, and Eran Zaidel's chapter gives a more balanced perspective. Sergent's work was stimulating and provocative, and has done much to sustain interest in cerebral asymmetry; this volume is a fitting memorial to her. □

Michael C. Corballis is in the Department of Psychology, University of Auckland, Private Bag 92019, Auckland, New Zealand.



THE new wave — an artist's impression of *Starwisp*, a hypothetical interstellar vehicle 'pushed' by microwave radiation. The future of space exploration is discussed in *Where Next Columbus?* edited by Valerie Neal. Essayists include Stephen Jay Gould, Timothy Ferris, Carl Sagan, Edward C. Stone, Robert L. Forward, Walter E. Massey and Thomas E. Lovejoy. Oxford University Press. \$35, £19.99.



Down in the mouth — a 'Goldschmidt' toad which, despite its remarkable development alteration, was found doing well in a garden in Hamilton, Canada. From the cover of *Phenotypes* by C. David Rollo. This academic tome "stresses the importance of genomic integration as a selectable feature in the evolution of multicellular phenotypes". Chapman and Hall, £45, \$85.

Remains of the day

Peter Andrews

Vertebrate Taphonomy. By R. Lee Lyman. Cambridge University Press: 1994. Pp. 524. £50, \$79.95 (hbk); £24.95, \$34.95 (pbk).

THE way in which animal bones become preserved and fossilized is a matter of increasing interest, not just for archaeologists and palaeontologists but also for those interested in the biochemistry and mineralogy of animal and plant remains. The study of fossilized remains, or 'taphonomy', covers all aspects of preservation from the time the remains were parts of living organisms to their incorporation in geological strata and removal for study in the laboratory. Taphonomy tells us about conditions of survival of the remains and is fundamental to biological and ecological interpretations of the fossil record (and therefore important to those who attempt to extract DNA from old bones or who study diagenetic processes). But the field has suffered from the lack of a comprehensive and authoritative treatment; Lyman has now provided this in *Vertebrate Taphonomy*.

The first three chapters provide a short history, with examples, of the study of taphonomy. The author takes pains to formalize the subject in terms of the scientific method and its implicit assumptions. He sees the principles of uniformitarianism and actualism as being fundamental to taphonomy and recognizes the progressive nature of taphonomic evidence. Implicit in his argu-

ment is the question of what information is provided by the study of taphonomy. On the one hand, taphonomy may be seen as a necessary precursor to reconstructions of past environments by first reconstructing past faunas; on the other, it may itself be viewed as an extra source of information about the environment, revealing aspects of environmental processes that might not be apparent from more traditional approaches.

The main part of the book contains detailed accounts of taphonomic processes. Chapters 4 to 7 deal with dispersal and accumulation of vertebrate remains, whereas chapters 8 to 11 consider surface and subsurface modifications. Chapter 4 combines an excellent summary of the vertebrate skeleton, including basic background on bone histology, with a useful account of some of the ways of counting vertebrate remains according, for example, to numbers of elements, animal units or individuals. This methodology is an obvious strength of the author's own work. There then follows a chapter on population parameters and disarticulation patterns of skeletons. This is an odd mixture: the latter topic flows on naturally from the previous chapter to the next two chapters on the processes of accumulation and dispersal and the frequencies of vertebrate remains; but the subject of population parameters is misplaced here — indeed, it could be argued that it has no place at all in this book, as the topic is more a matter of palaeoecological interpretation than taphonomic reconstruction.

The comprehensive treatment of surface modifications provides a valuable resource for future research. The effects

of all sorts of natural processes are documented in detail and there is an informative section on human modifications. The descriptions of changes to bones after burial are less authoritative, mainly because such modifications have been less well studied by vertebrate taphonomists. More disappointing is the chapter on diagenesis: much of the work by mineralogists is neither included nor referenced.

Vertebrate taphonomy has tended to focus on mammal remains, partly because many palaeontologists who study birds, fish, amphibians and reptiles have still to recognize the need to consider taphonomic processes. There is nevertheless a useful chapter that summarizes taphonomic research on these four groups of lower vertebrates. The methods have evidently been based on those devised for mammals, although it is questionable whether either the taphonomic processes themselves or their effects on bone histologies are the same.

Vertebrate Taphonomy concludes with an excellent discussion chapter, a marvellous bibliography running to 35 pages and a good glossary and index. They round off a comprehensive, detailed and accurate book that will make taphonomy accessible to specialists and nonspecialists alike. It is essential reading for all vertebrate palaeontologists, archaeologists and palaeoecologists and provides crucial background information for workers in other disciplines who use fossil material, especially molecular biologists attempting to extract DNA from fossil bone. □

Peter Andrews is at the Natural History Museum, Cromwell Road, London SW7 5BD, UK.

Atmospheric effects of the Mt Pinatubo eruption

M. Patrick McCormick, Larry W. Thomason & Charles R. Trepte

The eruption of Mt Pinatubo in June 1991 caused the largest perturbation this century to the particulate content of the stratosphere. The radiative influence of the injected particles put an end to several years of globally warm surface temperatures. At the same time, the combined effect of volcanic particles and anthropogenic reactive chlorine has led to record low levels of stratospheric ozone.

It is well known that human activity is perturbing the chemical composition and radiative balance of the Earth's atmosphere. Studies¹ of the sensitivity of our climate to increasing concentrations of greenhouse gases, so named for their ability to retain heat in the atmosphere, predict that the increase of CO₂ concentration from a pre-industrial value of ~270 p.p.m. to 600 p.p.m. by the middle of the next century, along with expected increases in other greenhouse gases, will increase global surface temperature by 2–5 °C. This picture is complicated by the increasing concentrations of anthropogenic aerosols in the lower troposphere, which act to mitigate greenhouse warming². In the stratosphere, chlorine concentrations have increased because of anthropogenic chlorofluorocarbon production³. Reactive chlorine compounds play an important role in the chemical processes that give rise to the Antarctic ozone hole, and the weight of evidence suggests that they also contribute to global losses of ozone, with possible concomitant increases in the intensity of biologically harmful ultraviolet-B radiation reaching the Earth's surface⁴.

The 15 June 1991 eruption of Mt Pinatubo (15° N, 121° E), on the island of Luzon in the Republic of the Philippines, showed that there is also significant natural, though episodic, variability in the composition of the atmosphere, particularly that of the stratosphere. The Mt Pinatubo eruption forced the evacuation of more than 200,000 people and caused the immediate deaths of more than 300, many from the collapse of homes due to the combination of heavy ashfall and rain from the nearly simultaneous passage of Typhoon Yunya. The threat to life and property from remobilization of tephra ashfall and pyroclastic material persisted through the remainder of 1991 and required careful monitoring by local authorities⁵.

The volcanic plume associated with the 15 June eruption was observed to have reached an altitude of more than 30 km. In addition to particulate matter, the eruption also injected gaseous SO₂ into the stratosphere⁶, which in turn was transformed into H₂SO₄/H₂O aerosol (particles suspended in air). With an estimated⁷ peak aerosol mass loading of 30×10^{12} g (30 Tg), the eruption of Mt Pinatubo caused what is believed to be the largest

aerosol perturbation to the stratosphere this century (Table 1, updated from ref. 8), but it is still smaller than the estimated aerosol perturbations from the eruptions of Tambora in 1815 (>100 Tg) and Krakatau in 1883 (~50 Tg)⁹. Given both recent volcanic history and inferences from ancient eruptions such as that of Toba (~1,000 Tg of SO₂) ~75,000 yr ago⁹, the 1991 Mt Pinatubo eruption is clearly not exceptional in its modification of the stratospheric aerosol loading. On the other hand, the Mt Pinatubo eruption is unique in the sense that it has been the most intensely observed eruption on record, as its volcanic cloud has been monitored by ground-based and aircraft lidars and solar photometers, various balloon-borne and airborne aerosol counters and other instruments, and many satellite-borne instruments. From a radiative perspective, the aerosol contributed to the end of several years of globally warm surface temperatures experienced in the late 1980s and early 1990s. The aerosol loading has also been associated with chemical and dynamical perturbations affecting the concentration of NO₂, reactive chlorine and ozone. Attempts to model the dispersion of the aerosol and the radiative and chemical impact of the eruption are 'acid tests' of our understanding of atmospheric processes (Fig. 1).

Mt Pinatubo aerosol and dynamical processes

Following the 15 June eruption, the evolving cloud of water vapour, sulphurous gases and aerosol moved westwards and circled the globe in approximately 22 days⁶. According to aircraft, ship, and ground-based lidar measurements, the bulk of Mt Pinatubo aerosol was initially concentrated between 20 and 27 km in altitude, although some was observed as low as the tropopause and as high as 30 km. The total mass of SO₂ injected by the eruption was measured¹⁰ by the total ozone mapping spectrometer (TOMS) to be approximately 20 Tg. This SO₂ was eventually converted into H₂SO₄/H₂O aerosol, beginning with an oxidative process involving OH[•] in a characteristic time⁶ of 30 days. The SO₂ mass estimate is consistent with the 30-Tg peak in total stratospheric aerosol mass loading estimated by the solar occultation instrument SAGE II (the stratospheric aerosol and gas experiment-II) in the last few months of 1991.

A notable feature of the Mt Pinatubo eruption was the rapid movement of a substantial fraction of volcanic material across the Equator to about 10° S during the first two-week period following the eruption. Within the next few weeks, the cloud continued to disperse longitudinally, but more slowly in the meridional direction, and occupied the latitude band of approximately 20° S to 30° N (Fig. 2)^{11–14}. The degree of early cross-equatorial transport was unusual because the early dispersion of volcanic material from other eruptions near the subtropics, such as El Chichon (17° N, 93° W), remained mostly in their hemisphere of origin. The amount of Mt Pinatubo volcanic material transported south at altitudes above 24 km was highlighted by aerosol observations by SAGE II in July and September and by the improved stratospheric and mesospheric sounder

TABLE 1 Major twentieth-century eruptions

Volcano	Date	Estimated aerosol loading (Tg)
Stratospheric background	Possibly 1979	<1
Katmai	June 1912	20
Agung	March 1963	16–30
Fuego	October 1974	3–6
El Chichon	April 1982	12
Mt Pinatubo	June 1991	30
Cerro Hudson	August 1991	3

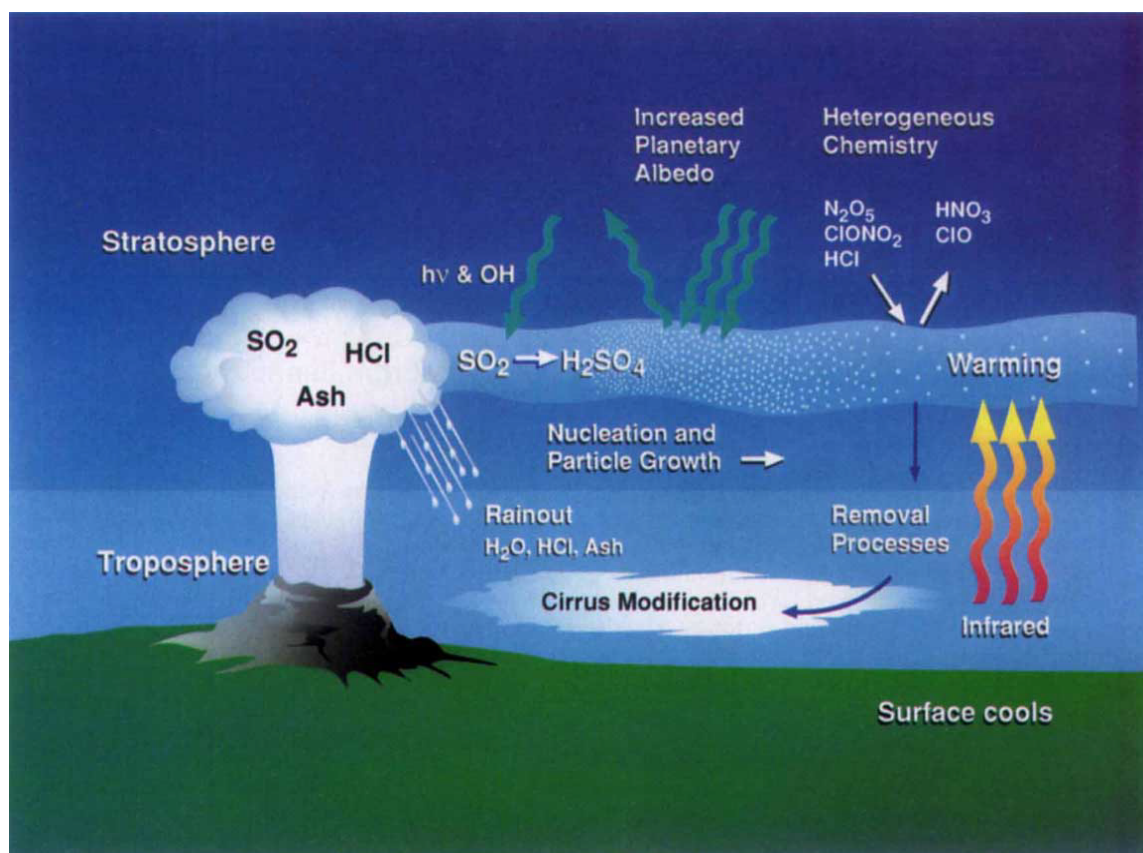


FIG. 1 As shown in this diagram, a volcanic eruption can produce a significant perturbation to the Earth-atmosphere system by injecting material into the stratosphere where, depending on the magnitude and altitude of the injection, it may persist for several years. The injected material may include ash, which typically does not remain for more than a few months, and gaseous components including water vapour, sulphur dioxide and hydrochloric acid. Most hydrochloric acid is dissolved into condensing water vapour and quickly rains out of the original cloud. Aerosols are produced when the sulphur dioxide (SO_2) is chemically transformed into sulphuric acid (H_2SO_4) which rapidly condenses into aerosols because it has a very low saturation vapour pressure. The new aerosol increases the Earth's albedo by reflecting solar radiation

back into space, and can warm the stratosphere by absorbing upwelling infrared radiation. Sedimentation and atmospheric circulation eventually transport the aerosol into the troposphere where it may modify cloud optical properties (particularly cirrus) and further modify the Earth's radiative processes. An additional effect of an eruption is the increased efficiency of heterogeneous chemical processes (that is, processes that take place on the surface of aerosols). This effect, coupled with anthropogenically increasing stratospheric chlorine levels, leads to ozone destruction by modifying the chemistry of reactive chlorine and nitrogen. This ozone removal process is similar to that which produces the Antarctic ozone hole, except that the surface in the latter case is provided by polar stratospheric clouds⁴.

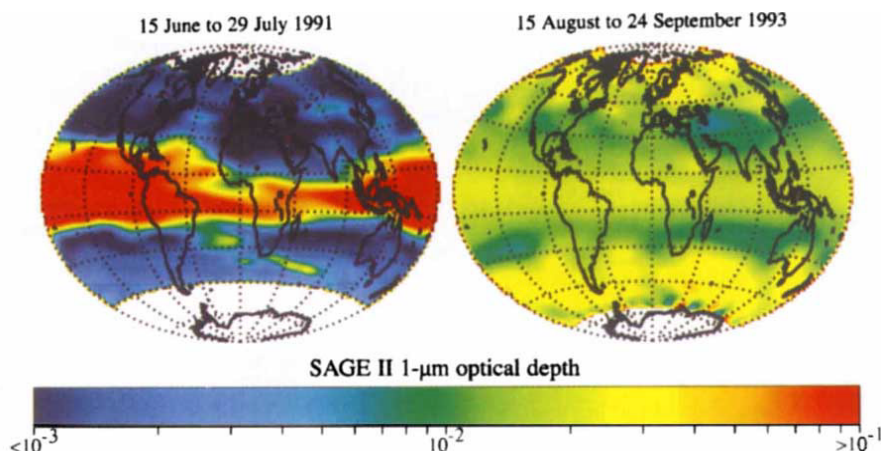
(ISAMS)¹⁵ in October which showed that maximum aerosol opacity resided near 10°S .

Recent numerical transport simulations suggest that the southward displacement of the aerosol plume within the tropics was the result of meridional circulations induced by local heating, rather than planetary-scale transport patterns¹⁶. The heating took place through the absorption of upwelling infrared radiation by the newly nucleated aerosol, and was responsible for warming the lower tropical stratosphere through the remainder of 1991. Peak temperatures at the 30-hPa level in the atmosphere ($\sim 24\text{ km}$) were found in September 1991 to be as much as three standard deviations above the 26-year monthly mean (increases of approximately 3.5 K)¹⁷. The positive temperature anomalies gradually decreased during 1992 as the tropical aerosol dispersed. Strong local heating not only affected horizontal transport, but also enhanced upward motions in the tropics. Satellite observations detected the lifting of aerosol to altitudes exceeding 35 km by October 1991. Because ozone concentration peaks near 25 km in the tropics, a more vigorous circulation at low latitudes, moreover, would have carried ozone-poor air to higher altitudes. At altitudes below 25 km , ozone is long-lived and a gradual photochemical relaxation to pre-eruption levels would have

occurred in about 1–3 months. This mechanism could be responsible for a reduction of columnar ozone of about 6–8% observed over the Equator during the first several months after the eruption^{18–21}.

One feature that seems to be characteristic of large eruptions at low latitudes is the relatively slow transport timescales of volcanic material out of the tropics. Observations following the eruptions of El Chichon in 1982 and Nevada del Ruiz in 1985 revealed the presence of enhanced reservoirs of volcanic aerosol at low latitudes which persisted for several years²². Poleward transport of aerosol into the winter hemisphere tends to be suppressed when easterly winds lie over the Equator²³. During these conditions, strong horizontal wind shear (velocity gradient) lies in the subtropics, separating the tropical easterlies and extratropical westerlies in the winter hemisphere. Wind shear inhibits the horizontal mixing of air. When westerly winds lie over the Equator, however, horizontal wind shear is weaker in the subtropics and meridional mixing occurs more readily. The reversal between easterly and westerly winds in the lower tropical stratosphere is known as quasi-biennial oscillation (QBO). It dominates the circulation in the lower tropical stratosphere and has a period of about 28 months.

FIG. 2 The development, dispersion, and dissipation of the aerosol from the Mt Pinatubo eruption was monitored by several instruments on satellites. Some interesting observations of the early confinement of the aerosol to the tropical stratosphere were provided by the advanced very high resolution radiometer (AVHRR)^{11,14}. Column atmospheric aerosol optical depth determined by AVHRR is usually dominated by boundary-layer aerosols (the lowest few kilometres of the atmosphere) but, after the Mt Pinatubo eruption, possessed a clear volcanic signature until at least early 1993¹⁴. Several instruments aboard the upper atmosphere research satellite (UARS) were adversely affected by the high concentration of aerosols after the eruption but were also able to provide interesting measurements of the aerosol. For instance, the improved stratospheric and mesospheric sounder (ISAMS) and the halogen occultation experiment (HALOE) made estimates of the relative $\text{H}_2\text{SO}_4\text{-H}_2\text{O}$ proportions using strong H_2SO_4 absorption features between 7 and 12 μm (ISAMS) and in the 2.45–5.26 μm wavelength band (HALOE)^{15,78}. The stratospheric aerosol and gas experiment (SAGE II), which has been functioning aboard the earth radiation budget satellite (ERBS) since 1984, has made stratospheric aerosol extinction profiles in the visible and near-infrared throughout this period. This figure shows integrated SAGE II stratospheric optical depth for two periods after the eruption of Mt Pinatubo. The data from 15 June to 29 July 1991 (left) show the tropical confinement of the aerosol and the



increase (approximately two orders of magnitude) in the 1- μm (actually 1.020- μm) optical depth⁷. Some indications of the initial transport to middle and high latitudes are also evident. SAGE II measurements indicated that by early 1992, the stratospheric optical depth was at, or exceeded, 0.1 at all latitudes. The data for 15 August to 24 September 1993 (right) shows the result of the dispersal and gradual removal of aerosol from the stratosphere. Although the stratospheric optical depth has decreased by approximately one order of magnitude by this time, it is also evident that this optical depth was still dominated by volcanically derived aerosol.

A similar signature of relatively slow tropical aerosol transport to high latitudes was evident after the eruption of Mt Pinatubo. Easterly winds prevailed above about 23 km in the lower tropical stratosphere for the first 6 months after the eruption. Above this height, airborne lidar^{24,25} and satellite²⁶ observations of the volcanic plume found steep meridional gradients of aerosol near 20° N and 20° S. At lower altitudes, however, volcanic aerosol dispersed more quickly towards the poles, and it was primarily this low-altitude transport that was responsible for the initial lidar observations of the Pinatubo aerosol (and the spectacular sunsets) in northern mid-latitudes shortly after the eruption^{27–29}.

Over the course of time, volcanic aerosol is transported from the stratosphere into the troposphere where a variety of mechanisms (mostly associated with clouds and precipitation) effect the deposition of aerosol to the Earth's surface. Gravitational sedimentation of the aerosol, subsidence (downward air motions) in the winter hemisphere, and episodic transport of stratospheric air into the troposphere through tropopause folds contribute to the removal of aerosol from the stratosphere. By the end of 1993, about 2½ years after the eruption, the global stratospheric aerosol mass loading had decreased to approximately 5 Tg. Since the middle of 1992, the total stratospheric aerosol mass has decreased with a 1/e-folding time of approximately 1 year. Measurements by lidars, Sun photometers, and satellite instruments (for example, SAGE II), showed that the longitudinally averaged stratospheric opacity at a wavelength of 1 μm had declined from a maximum in excess of 0.2 in the tropics in late 1991 to a range over the globe of 0.006–0.03 by mid-1993 (Fig. 2). These values still represent enhancement by a factor of 5–10 of values observed in 1989 and 1990, when the global average stratospheric opacity at 1 μm was between 0.001 and 0.003.

Effect on the Earth's radiative processes

The introduction of large amounts of sulphuric acid aerosol into the stratosphere increases the planetary albedo (essentially the Earth's reflectivity of solar radiation) because these aerosol particles are efficient scatterers but only weak absorbers at solar wavelengths. An analysis of data from the Earth radiation budget experiment (ERBE) aboard the Earth radiation budget

satellite (ERBS) found that the albedo in July, August and September 1991 increased compared to a five-year mean³⁰. The global albedo in August 1991, for example, was 0.250, or more than five standard deviations greater than the five-year mean of 0.236. The greatest effect occurred over cloud-free regions, where the normally low albedo increased by more than 20%. An increase in albedo was also noted in the normally high-albedo regions associated with deep convective cloud systems. This may represent an indirect effect of the volcanic aerosols in the sense that aerosol particles transported into the upper troposphere are incorporated into deeper convective clouds and alter their microphysical structure³¹. Increasing the number of aerosol particles available to act as cloud condensation and ice nuclei tends to reduce the mean size of the resulting cloud particles. Because smaller particles are more efficient scatterers of visible radiation, this process increases the albedo of the cloud³². It has also been suggested that optically thin (but radiatively important) cirrus may be subject to similar microphysical changes^{32,33}, although it is not at present clear if any changes in the optical properties of cirrus have occurred. The changes in the Earth's albedo observed by ERBE resulted in a net cooling of approximately 8 W m⁻² between 5° S and 5° N, with a net cooling of 4.3 W m⁻² between 40° S and 40° N. The meridional distribution of the cooling and its magnitude in the tropics reflects the confinement of the aerosol as discussed earlier.

To a first approximation, an increase in stratospheric opacity, such as that following the eruption of Mt Pinatubo, should cool the Earth because the aerosol increases the albedo of the Earth-atmosphere system and therefore increases the amount of solar radiation scattered back into space. But microphysical properties of the aerosol (in particular, aerosol size) determine whether the infrared absorptivity of stratospheric aerosols, which acts to warm the Earth (that is, the greenhouse effect), dominates the scattering of incoming solar radiation back into space (that is, the albedo effect), which acts to cool the Earth. It has been shown³⁴ that if the aerosol effective radius (the area-weighted mean radius of the aerosol size distribution) exceeds ~2 μm , the warming effect overrides the cooling effect. Estimates of effective aerosol radius following the Pinatubo eruption indicate that an increase from an average of 0.3 μm before the

eruption to about 1 μm afterwards occurred, and thus cooling would be expected^{35–37}.

As previously noted, substantial heating in the stratosphere was observed immediately after the eruption. This heating was sufficient to cause tropical stratospheric temperatures to increase as much as three standard deviations above the 26-year mean at 30 hPa (~24 km). Analyses by Christy and Spencer³⁸ using the satellite-borne microwave sounding unit (MSU) also indicate warmer than average stratospheric temperatures in 1991 and 1992. By the end of 1993, however, stratospheric temperatures had decreased to the lowest values ever observed by MSU. This cooling may be related to the ozone loss observed in the stratosphere (to be discussed later), as ozone is an effective absorber of solar radiation and is responsible for substantial heating in the stratosphere.

The negative radiative forcing associated with the eruption of Mt Pinatubo exceeded the magnitude of the positive forcing associated with the greenhouse gases during the second half of 1991 and much of 1992, and remained significant through 1993 (Fig. 3). In response to this forcing, the mean tropospheric temperature was 0.2 °C below normal in 1992 compared with a base period of 1958–91³⁹. The 1992 El Niño–Southern Oscillation (ENSO, a feature of Pacific Ocean circulation with important couplings to tropospheric weather patterns) would have caused a warmer than average troposphere. As a result, the 1992 tropospheric temperature anomaly, adjusted for ENSO⁴⁰, was –0.4 °C, a decrease of more than 0.7 °C from 1991. Data

from MSU yield a similar global decline in lower tropospheric temperatures of 0.5 °C in mid-1992, with much of the decrease occurring in the Northern Hemisphere (0.7 °C)⁴¹. In a global-average context, the measured temperature anomaly is consistent with the value of –0.5 °C estimated from a global climate model⁴². It should be noted that the small mean changes in global temperature are not evenly distributed, but rather consist of regions of both relative cooling and warming whose locations and magnitudes are dependent on season⁴². The temperature anomaly in 1993 was smaller than in 1992 with a value of –0.1 °C, which, adjusted for the phase of the ENSO, is equivalent to an anomaly of –0.2 °C (J. Angell, personal communication).

This drop in global mean temperatures is similar to that observed in 1964 which had the largest anomaly in the 1958–92 analysis period. Significantly, 1964 is the year following the eruption of Agung, which appears to be the second largest volcanic eruption (after Pinatubo) to have occurred since 1958. The departure is also comparable to the estimated impact of Tambora (–0.4 to –0.7 °C) and greater than that estimated for El Chichon (–0.2 °C) and Krakatau (–0.3 °C)⁹. That the Pinatubo cooling exceeds that of Krakatau is surprising because the latter is believed to have put more SO₂ into the stratosphere. However, this anomaly may simply reflect uncertainties in the estimates of both stratospheric aerosol loading and global surface temperature, particularly for the earlier eruption.

Effect on stratospheric chemical processes

Observations of the input of hydrochloric acid (HCl) soon after the eruption of Mt Pinatubo showed little or no increase compared to pre-eruption levels^{43,44}. The measurements are consistent with modelling efforts by Tabazadeh and Turco⁴⁵ that suggest that the bulk of erupted chlorine is removed within the volcanic plume by condensation into supercooled water drops which are then rained out. These findings argue strongly that stratospheric chlorine is primarily a product of anthropogenic chlorofluorocarbons rather than volcanic eruptions⁴⁶.

The concentrations of chlorine species, and particularly reactive chlorine species such as Cl⁺ within the stratosphere are crucial in determining the rate of ozone destruction. Ozone loss associated with reactive chlorine figures prominently in the formation of the Antarctic ozone hole. In the Antarctic, the extreme low temperatures of the winter and early spring permit the formation of polar stratospheric clouds (PSCs)⁴⁷. The cloud particles provide sites for heterogeneous chemical reactions (that occur on the surface of an aerosol or cloud particle) which transform relatively inert forms of chlorine such as HCl and ClONO₂ into more reactive forms such as Cl₂. The rates of such reactions are strongly dependent on the total particulate surface-area density. When the chemically modified air is transported into sunlit regions (usually in late August and September at high southern latitudes) Cl₂ disassociates into highly reactive Cl⁺ which in turn is responsible for ozone destruction via gas-phase chemical reactions⁴.

An additional factor governing the depletion in the ozone hole is that the initial stage of PSC formation involves sequestration of a significant fraction of stratospheric nitric acid (HNO₃) in condensed forms. Because HNO₃ is unreactive, PSCs act as a sink for more reactive forms of nitrogen such as NO₂ and N₂O₅. Reactive nitrogen is a buffer to ozone loss because it is involved in reactions which convert reactive chlorine into non-reactive forms. There is abundant evidence that further growth of PSC particles by the deposition of ice followed by sedimentation of the larger particles leads to an irreversible denitrification (and dehydration) of the polar stratosphere, further enhancing ozone loss^{48–50}. The lack of an ozone hole in the Arctic (so far) can be attributed at least in part to generally warmer winter temperatures and, as a result, fewer PSC occurrences⁴, less denitrification and a shorter season.

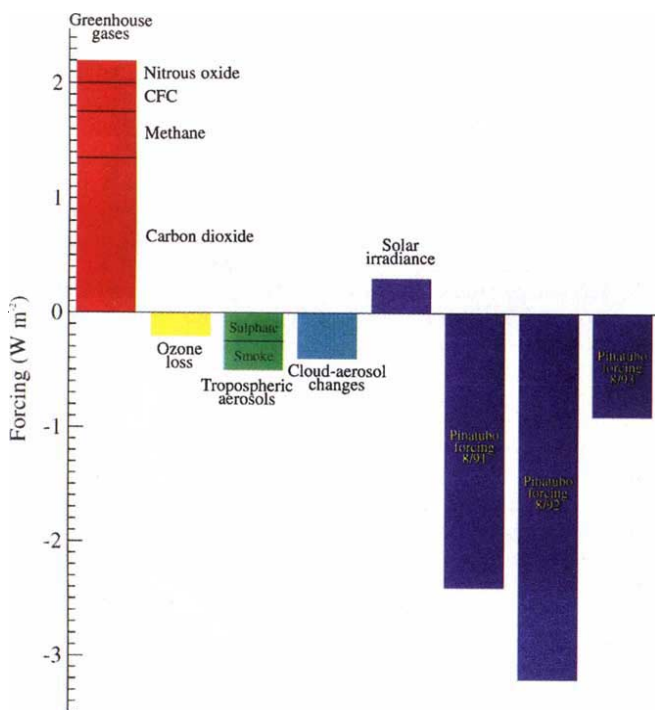


FIG. 3 Pre-industrial-to-modern global climate forcing, 1850–1990 (histogram adapted from Hansen *et al.*⁷⁹). Climate forcing is the change in the Earth's radiative balance resulting from either natural or anthropogenic sources. The largest forcing (~2 W m^{–2}) is due to the increase in greenhouse gases associated with industrialization, and the net forcing in this period remains positive when several smaller (and negative) effects, such as increasing tropospheric opacity due to human activities and global ozone loss, are included. Mt Pinatubo is estimated to have produced a transient negative forcing which exceeded greenhouse-gas forcing between mid-1991 and the end of 1992. It also made 1992 one of the coolest of the past 30 years. As the effect of Mt Pinatubo continues to decrease, the increasing magnitude of greenhouse-gas forcing should reveal itself in a return to the globally warm years observed in much of the 1980s.

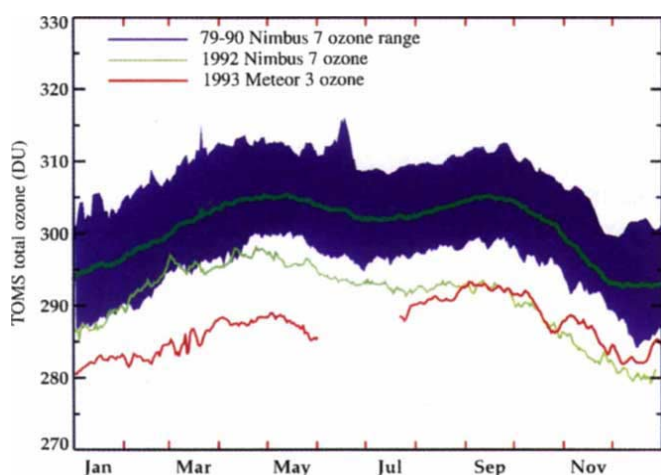


FIG. 4 Global mean ozone in Dobson units (DU) from the total ozone mapping spectrometer (TOMS) measurements as a function of time. Light green line, TOMS/NIMBUS 7 measurements for 1992; red line, TOMS/METEOR 3 measurements for 1993. These two lines are superimposed on the range of TOMS-observed global (65° N–65° S) mean ozone between 1979 and 1990 (dark blue line). The dark green line is the average total ozone for this period. TOMS data show a notable decrease in column ozone amounts beginning in early 1992 and persisting to the end of 1993. The loss reached as much as 6% of the column mean in April 1992^{72,73}.

After a large eruption such as that of Mt Pinatubo, stratospheric aerosol surface-area density may be large enough for heterogeneous reactions to perturb the concentrations of relevant chemical species not only in polar latitudes but also in middle and tropical latitudes. For instance, ozone loss (although much smaller than that reported below associated with the Mt Pinatubo eruption) attributed to increased aerosol loading was reported following the eruption of El Chichon in 1982⁵¹. Because elevated chlorine levels in the stratosphere are the result of chlorofluorocarbon emissions, this is an interesting example of the coupling of human and natural perturbations of stratospheric composition that lead to additional perturbations, in this case ozone destruction⁵². It has also been noted that the increase in important greenhouse gases like CO₂, CH₄ and N₂O (as well as CO) have slowed or stopped in the aftermath of the Pinatubo eruption⁵³. The cause for these changes is still under investigation.

An early indicator during the post-Pinatubo period of a connection between aerosols and ozone loss was found in association with aerosol formed by the eruption of Cerro Hudson (46° S; August 1991) rather than Mt Pinatubo. Stratospheric aerosol from Cerro Hudson was concentrated at low altitude (<16 km) and was able to penetrate deeply into high southern latitudes soon after that eruption. At that time, Pinatubo aerosol in the Southern Hemisphere was primarily residing above 16 km, with the impedance to meridional transport associated with the polar vortex boundary effectively blocking the aerosol from high latitudes until the vortex dissipated in November 1991⁵⁴. Before 1991, Antarctic ozone loss had not been observed below the base of the vortex (~14 km). But in the densest portions of the Cerro Hudson aerosol layer (11–13 km), ozone decreased by 50% in the 30 days following the arrival of the aerosol over Antarctica⁵⁵. The aerosol surface-area density in the layer was as high as 100 μm² cm⁻³, or 20–30 times higher than observed in 1990. Aerosol levels at these altitudes remained unusually high during the austral springs of 1992 and 1993 because of the downward transport of Pinatubo aerosol within the Antarctic vortex^{50,56}. Ozone levels in the 12–14 km range were extremely low in 1992

and 1993 with near-zero values in mid-October⁵⁷ in both years. Record low column ozone was also observed in these years (for example, 105 Dobson units (DU) on 11 October 1992 and 91 DU on 12 October 1993; column ozone amounts of less than 100 DU were noted on several occasions in 1993⁵⁷).

In the austral autumn of 1992, more evidence of unusual heterogeneous processing was observed over Antarctica. Elevated levels of chlorine dioxide (OCIO) were detected over McMurdo station as early as 13 April⁵⁸. Before 1992, elevated OCIO was observed only during winter and spring (June–September) as a by-product of heterogeneous chemical reactions on the surface of PSCs. The OCIO detected early in 1992 and the ozone losses in 1991 may be the by-products of heterogeneous reactions on the surface of volcanic aerosols:



where the moderating influence represented by the photolysis of HNO₃ back to NO₂ (allowing benign ClONO₂ to form)⁵⁹ was suppressed by the low solar illumination of Antarctica in the autumn. These findings suggest that there was considerable potential for ozone destruction by reactive chlorine in the Antarctic during times when it was too warm for PSCs to occur⁵⁸.

On the other hand, evidence of global-scale heterogeneous processing following the eruption of Mt Pinatubo was noted as early as August 1991, when measurements in New Zealand⁶⁰ showed a reduction in the column NO₂ that reached as much as 40% in October 1991. Significant column NO₂ decreases have also been reported in northern mid-latitudes⁶¹. Similar, although smaller, decreases in the column NO₂ were also reported after the eruption of El Chichon in 1982^{51,62}. SAGE II measurements showed that the loss of NO₂ occurred globally, largely at altitudes below 25 km and exceeded 50% at some altitudes. Stratospheric NO₂ concentrations remained suppressed throughout 1992 before returning to normal ranges in late 1993. The decreases in NO₂ suggest a repartitioning of reactive nitrogen species in favour of nitric acid (HNO₃) via the heterogeneous reaction (1)⁶³, a conclusion supported by recent reports of elevated gaseous HNO₃ at low latitudes⁶⁴ and mid-latitudes^{65–67}.

As previously mentioned, column ozone in the tropics decreased by 6–8% in the months following the Pinatubo eruption. The bulk of the loss was observed below 28 km and was as large as 20% near 24–25 km (refs 18, 68, 69). Also, small increases were noted above 30 km. Ozone losses in the lower tropical stratosphere were probably produced by increased lofting associated with stratospheric heating, associated with the new aerosol. In this scenario, ozone is lofted to altitudes at which photochemical processes are more effective in destroying it^{19,20}. Other processes may also have contributed to the ozone perturbations in the tropics, including those involving SO₂. Before conversion to sulphuric acid aerosol, SO₂ absorbs ultraviolet radiation thereby reducing the photolysis of O₂ and thus the production of ozone, particularly below 25 km. Above 25 km, photolysis of SO₂ contributes to chemical processes which produce⁷⁰ O₃. It has also been suggested that increased scattering of solar radiation by volcanic aerosol may perturb the ozone balance⁷¹.

Six months after the eruption of Mt Pinatubo, global mean column ozone, as measured by TOMS and shown in Fig. 4, began to show a significant downward trend with respect to the mean 1979–90 annual ozone cycle. By mid-1992, the global mean column ozone amount was significantly less than the minimum values measured between 1979 and 1990⁷². This downward trend continued well into 1993; in April of that year a deficit of approximately 6% (~18 DU) was observed, nearly 4% less than ever observed by TOMS before the eruption^{72,73}. A slow recovery toward pre-eruption values began in the second half of 1993. It should be noted, however, that global values are somewhat

deceptive because TOMS observed much larger decreases in the Northern Hemisphere than in the Southern Hemisphere⁷³. Dobson spectrophotometer observations from several sites in the United States showed a downward trend beginning in May 1992 which accelerated in the first half of 1993, establishing record low values between January and April of that year⁷⁴. On average, these sites measured a decrease of more than 10% compared to the 1979–91 monthly means. Departures were noted throughout 1992 and 1993, although the decrease was smaller in magnitude during May–August. Ozone-sonde data from Boulder, Colorado indicated⁷⁵ that this loss was greatest between 13 and 22 km altitude and averaged about 20%. Small increases in ozone concentration were noted above 25 km, though it should be noted that the total column amount is small above that altitude. The altitude of the most significant losses was highly correlated with the altitude of highest aerosol loading.

The observation of highly perturbed values of NO₂ and HNO₃, and the general correlation of the perturbations with enhanced aerosol, strongly indicates the importance of heterogeneous chemical processes in the destruction of ozone. Many of the likely processes are similar to those associated with ozone loss in the Antarctic, in which a loss of reactive nitrogen to non-reactive nitric acid led to enhanced levels of reactive chlorine. The reactions that produce increased reactive chlorine and thus increase ozone destruction include⁷⁶ reactions (1) and (2), and other reactions involving ClONO₂, HCl and HOCl. Ozone destruction reaches a maximum in the winter and early spring because the direct chlorine reactions are strongly temperature dependent (progressing more rapidly with decreasing temperature), and the photolysis of HNO₃ back to reactive nitrogen (which moderates ozone destruction by reactive chlorine in the lower stratosphere) is inhibited by the low solar illumination. Hanson *et al.*⁷⁶ also pointed out that these reactions occur slowly, and that long periods under favourable conditions were

required for significant ozone destruction. In the tropics, where solar illumination is high, little ozone loss is noted beyond mid-1992 despite the presence of very high aerosol loading. Modelling efforts by Hanson *et al.*⁷⁶ and Rodriguez *et al.*⁷⁷ corroborate the importance of heterogeneous reactions in the observed ozone loss.

Looking to the future

The eruption of Mt Pinatubo was a disaster for those living nearby, but for those trying to understand the natural and anthropogenic processes for global change, the eruption represented perhaps a once in a lifetime opportunity. Observations of the Mt Pinatubo aerosol, various chemical species, and the radiative consequences during both the most heavily loaded period and the subsequent recovery provide data needed to test dynamic, chemical and radiative modelling of the atmosphere. Verification of our understanding of the Earth–atmosphere climate system is crucial because future forecasts by such models affect national and international policies and regulations that will in turn directly affect our economic livelihood and quality of life. The Montreal Protocol and its subsequent modifications governing the production of chlorofluorocarbons are good examples of such actions. Of course, much more work is in progress about the effects of Mt Pinatubo than is discussed here. It is also clear that much more work is needed to exploit fully the opportunity provided by nature to enhance our knowledge of how this complex atmosphere and climate system works, and to provide us with a glimpse into the future. □

M. P. McCormick and L. W. Thomason are at the Atmospheric Sciences Division, NASA Langley Research Center, Hampton, Virginia 23681–0001, USA. Charles R. Trepte is at Science Applications International Corporation, Hampton, Virginia 23666, USA.

- Cess, R. D. *et al.* *Science* **245**, 513–516 (1989).
- Charlson, R. J. *et al.* *Science* **255**, 423–430 (1992).
- Gunson, M. R. *et al.* *Geophys. Res. Lett.* **21**, 2223–2226 (1994).
- Solomon, S. *Nature* **347**, 347–354 (1990).
- Pinatubo Volcano Observatory *Earth in Space* **4**, 5–10 (1992).
- Bluth, G. J. S., Doiron, S. D., Schnetzler, C. C., Krueger, A. J. & Walter, L. S. *Geophys. Res. Lett.* **19**, 151–154 (1992).
- McCormick, M. P. & Veiga, R. E. *Geophys. Res. Lett.* **19**, 155–158 (1992).
- Kent, G. S. & McCormick, M. P. *Optics News* **14**, 11–19 (1988).
- Pinto, J. R., Turco, R. P. & Toon, O. B. *J. geophys. Res.* **94**, 11165–11174 (1989).
- Bluth, G. J. S. *Nature* **327**, 386 (1993).
- Stowe, L. L., Carey, R. M. & Pellegrino, P. P. *Geophys. Res. Lett.* **19**, 159–162 (1992).
- DeFoor, T. E., Robinson, E. & Ryan, S. *Geophys. Res. Lett.* **19**, 187–190 (1992).
- Winker, D. M. & Osborn, M. T. *Geophys. Res. Lett.* **19**, 167–170 (1992).
- Long, C. S. & Stowe, L. L. *Geophys. Res. Lett.* **21**, 2215–2218 (1994).
- Grainger, R. G. *et al.* *Geophys. Res. Lett.* **20**, 1283–1286 (1993).
- Young, R. E., Houben, B. & Toon, O. B. *Geophys. Res. Lett.* **21**, 369–372 (1994).
- Labitzke, K. & McCormick, M. P. *Geophys. Res. Lett.* **19**, 207–210 (1992).
- Grant, W. B. *et al.* *Geophys. Res. Lett.* **19**, 1109–1112 (1992).
- Kinne, S. *et al.* *Geophys. Res. Lett.* **19**, 1927–1930 (1992).
- Schoeberl, M. R. *et al.* *Geophys. Res. Lett.* **20**, 29–32 (1993).
- Brasseur, G. & Granier, C. *Science* **257**, 1239–1242 (1992).
- Yue, G. K., McCormick, M. P. & Chiu, E.-W. *J. geophys. Res.* **96**, 5209–5219 (1991).
- Trepte, C. R. & Hitchman, M. H. *Nature* **355**, 626–628 (1992).
- Browell, E. V. *et al.* *Science* **261**, 1155–1158 (1993).
- Winker, D. M. & Osborn, M. T. *EOS (abstr.)* **73**, 625 (1992).
- Trepte, C. R., Veiga, R. E. & McCormick, M. P. *J. geophys. Res.* **98**, 18563–18573 (1993).
- Osborn, M. T., Winker, D. M., Woods, D. C., DeCoursey, R. J. & Trepte, C. R. *Proc. 17th Int. Laser Radar Conf.*, 446–449 (Sendai, Japan, July 25–29, 1994).
- Hayashida, S. *Global Volcanism Network Bull.* Vol. 6, sect. 16, 14–16 (Smithsonian Institution, Washington DC, 1991).
- Jager, H. *Geophys. Res. Lett.* **19**, 191–194 (1992).
- Minnis, P. *et al.* *Science* **259**, 1411–1415 (1993).
- Twomey, S. A., Piepgrass, M. & Wolfe, T. *Tellus* **36B**, 356–366 (1984).
- Sassen, K. *Science* **257**, 516–519 (1992).
- Jensen, E. J. & Toon, O. B. *Geophys. Res. Lett.* **19**, 1759–1762 (1992).
- Lacis, A., Hansen, J. E. & Sato, M. *Geophys. Res. Lett.* **19**, 1607–1610 (1992).
- Russell, P. *et al.* *J. geophys. Res.* **98**, 22969–22986 (1993).
- Valero, F. P. J. & Pilewskie, P. *Geophys. Res. Lett.* **19**, 163–166 (1992).
- Brock, C. A. *et al.* *Geophys. Res. Lett.* **20**, 2555–2558 (1993).
- Monastersky, R. *Science News* **145**, 70 (1994).
- Halpert, M. S. *et al.* *Eos* **74**, 433 (September 21, 1993, American Geophysical Union EOS Transactions).
- Angell, J. *Geophys. Res. Lett.* **17**, 1093–1096 (1990).
- Dutton, E. G. & Christy, J. R. *Geophys. Res. Lett.* **19**, 2313–2316 (1992).
- Hansen, J. E., Lacis, A., Ruedy, R. & Sato, M. *Geophys. Res. Lett.* **19**, 215–218 (1992).
- Mankin, W. G., Coffey, M. T. & Goldman, A. *Geophys. Res. Lett.* **19**, 179–182 (1992).
- Wallace, L. & Livingston, W. *Geophys. Res. Lett.* **19**, 1209 (1992).
- Tabazadeh, A. & Turco, R. P. *Science* **260**, 1082–1086 (1993).
- Taubes, G. *Science* **260**, 1580–1583 (1993).
- McCormick, M. P., Steele, H. M., Hamill, P., Chu, W. P. & Swisler, T. J. *J. Atmos. Sci.* **39**, 1387–1397 (1982).
- Fahey, D. W. *et al.* *J. geophys. Res.* **94**, 11299–11316 (1989).
- Kelly, K. K. *et al.* *J. geophys. Res.* **94**, 11317–11358 (1989).
- Thomason, L. W. & Poole, L. R. *J. geophys. Res.* **98**, 23003–23012 (1993).
- Hofmann, D. J. & Solomon, S. *J. geophys. Res.* **94**, 5029–5041 (1989).
- Rodriguez, J. M., Ko, M. K. W. & Sze, N. D. *Nature* **352**, 134–137 (1991).
- Kerr, R. A. *Science* **263**, 1562 (1994).
- Pitts, M. C. & Thomason, L. W. *Geophys. Res. Lett.* **20**, 2451–2454 (1993).
- Deshler, T. *et al.* *Geophys. Res. Lett.* **19**, 1819–1822 (1992).
- Kent, G. S., Trepte, C. R., Farrukh, U. O. & McCormick, M. P. *J. Atmos. Sci.* **14**, 1536–1551 (1985).
- Hofmann, D. J., Oltman, S. J., Lathrop, J. A., Harris, J. M. & Vömel, H. *Geophys. Res. Lett.* **21**, 421–424 (1994).
- Solomon, S., Sanders, R. W., Garcia, R. R. & Keys, J. G. *Nature* **363**, 245–248 (1993).
- Fahey, D. W. *et al.* *Nature* **363**, 509–514 (1993).
- Johnston, P. V., McKenzie, R. L., Keys, J. G. & Matthews, W. A. *Geophys. Res. Lett.* **19**, 211–213 (1992).
- Koike, M., Kondo, Y., Matthews, W. A., Johnston, P. V. & Yamazaki, K. *Geophys. Res. Lett.* **20**, 1975–1978 (1993).
- Johnston, P. V. & McKenzie, R. L. *J. geophys. Res.* **94**, 15499–15511 (1989).
- Kawa, S. R. *et al.* *Geophys. Res. Lett.* **20**, 2507–2510 (1993).
- Rinsland, C. P. *et al.* *J. geophys. Res.* **99**, 8213–8219 (1994).
- Koike, M. *et al.* *Geophys. Res. Lett.* **21**, 397–400 (1994).
- Mills, M. J. *et al.* *Geophys. Res. Lett.* **20**, 1187–1190 (1993).
- Webster, C. R., May, R. D., Allen, M., Jaeglé, L. & McCormick, M. P. *Geophys. Res. Lett.* **21**, 53–56 (1994).
- Hofmann, D. J. *et al.* *Geophys. Res. Lett.* **20**, 1555–1558 (1993).
- Hofmann, D. J. *et al.* *Geophys. Res. Lett.* **21**, 65–68 (1993).
- Bekki, S., Tsumi, R. & Pyle, J. A. *Nature* **362**, 331–333 (1993).
- Michelangelo, D. V., Allen, M. & Yung, Y. L. *J. geophys. Res.* **94**, 18429–18443 (1989).
- Gleason, J. F. *et al.* *Science* **260**, 523–526 (1993).
- Herman, J. R. & Larko, D. J. *geophys. Res.* **99**, 3483–3496 (1994).
- Komhyr, W. D. *et al.* *Geophys. Res. Lett.* **21**, 201–204 (1993).
- Hofmann, D. J. *et al.* *Geophys. Res. Lett.* **21**, 65–68 (1994).
- Hanson, D. R., Ravishankara, A. R. & Solomon, S. *J. geophys. Res.* **99**, 3615–3629 (1994).
- Rodriguez, J. M. *et al.* *Geophys. Res. Lett.* **21**, 209–212 (1994).
- Hervig, M. E., Russell, J. M., Gordley, L. R., Park, J. H. & Drayson, S. R. *Geophys. Res. Lett.* **20**, 1291–1294 (1993).
- Hansen, J. E., Lacis, A., Ruedy, R., Sato, M. & Wilson, H. *Res. Explor.* **9**, 143–158 (1993).

ACKNOWLEDGEMENTS. We thank J. Herman for providing the TOMS ozone trends, and L. Poole, G. Kent, J. Hansen and S. Solomon for helpful comments and suggestions.

Photodesorption from low-temperature water ice in interstellar and circumsolar grains

M. S. Westley*, R. A. Baragiola*, R. E. Johnson* & G. A. Baratta†

* University of Virginia, Thornton Hall, Charlottesville, Virginia 22901, USA

† Osservatorio Astrofisico, Università di Catania, Italy

DUST grains in the interstellar medium¹ and the outer Solar System²⁻⁴ commonly have a coating of water ice, which affects their optical properties and surface chemistry. The thickness of these icy mantles may be determined in part by the extent of photodesorption (photosputtering) by background ultraviolet radiation. But this process is poorly understood, with theoretical estimates of the photodesorption rate spanning several orders of magnitude^{5,6}. Here we report measurements of the absolute ultraviolet photodesorption yield of low-temperature water ice. Our results indicate that the rate of photodesorption is appreciable. In particular, it can account for the absence of icy mantles on grains in diffuse interstellar clouds, it exceeds solar-wind ion erosion and sublimation in the outer Solar System, and it is important in determining the lifetimes of icy mantles in dense molecular clouds.

The experiments were performed in an ultra-high-vacuum chamber (base pressure $\sim 1 \times 10^{-10}$ torr) using a quartz crystal resonator microbalance (sensitivity 0.1 monolayers of water)⁷ and a quadrupole mass spectrometer. Ice films were grown by vapour deposition on the cooled gold electrode of the microbalance⁷ to a thickness ($\sim 5,000$ Å) much larger than the attenuation depth of the 10.2-eV Lyman- α photons (~ 450 Å)⁸. Changing growth conditions to produce amorphous or crystalline cubic ice⁷ produced no detectable effect on photodesorption. Irradiation was done at normal incidence with light from a hydrogen microwave discharge which was shown to have negligible ultraviolet radiation outside the Lyman- α region. Calibration of the optical spectrometer and photon detectors were done at the SURF-II metrology line (NIST, Gaithersburg, Maryland).

Once an ice film is grown, its mass remains essentially stable until it is exposed to the ultraviolet radiation. When the ultraviolet source is turned on, the mass of ice decreases steadily; from this mass loss and the measured photon dose, we calculate the desorption yield Y as the number of molecules desorbed per incident photon. Mass spectrometry shows that the desorption flux is mainly H_2O , as expected⁹, with small amounts of H_2 and O_2 . The mass loss at large doses is proportional to the photon flux (Fig. 1), showing that desorption is not due to sublimation caused by absorbed heat. Furthermore, the mass loss dropped to unmeasurable levels when we used a window that blocked the Lyman- α radiation, but let the visible and infrared pass unattenuated to the sample.

Figure 2 shows the dependence of Y on photon dose and ice temperature. The low-dose behaviour shows that photodesorption does not occur because of absorption of a single photon, as assumed previously^{2,10}, but requires that the ice be altered by previous photoabsorption events. This result can be related to those for fast, light ions where the desorption yield at low temperatures and low excitation densities is proportional to the square of the excitation density^{11,12}. The observed photon-dose dependence of Y at low doses confirms that single excitation events, like those leading to direct dissociation or dissociative attachment, do not produce significant desorption in pure water ice.

A mechanism for desorption can involve the formation of precursors (probably radicals) which directly absorb the radiation or react with a nearby, newly excited molecule, converting potential energy into kinetic energy of the reaction products. Thus the incubation stage which we observe at low temperatures, before we can measure significant photodesorption, may result from the need to achieve a sufficient density of radical precursors in the surface layers of the ice films, so that the formation of a new photolysed radical is likely to lead to a desorption event. The incubation dose of $\sim 10^{18}$ photons cm^{-2} at 35 K decreased by a factor of about two in experiments in which a fresh 200-Å layer was deposited on a previously ultraviolet-saturated sample and then re-irradiated. Figure 2 also shows that incubation is shorter at higher temperatures, where radicals diffuse more easily¹³ and reactions are enhanced.

Figure 3 shows that the temperature dependence of the saturation photodesorption yield is similar to that of desorption yields under ion^{11,12} and electron¹⁴ bombardment. The temperature dependence of the yield is fitted by an expression $Y(T) = Y_0 + Y_1 e^{-E/kT}$ where T is temperature, with an activation energy $E \sim 0.03$ eV. This value is in the range associated with hydrogen diffusion¹⁵, and the sharp increase of Y occurs at temperature where OH becomes mobile¹⁵. The accumulation of radicals in the ultraviolet-irradiated ice was confirmed by thermal desorption experiments. Mass spectrometry of the gas produced by heating the ice to 160 K at ~ 1 K min^{-1} showed peaks at mass 33 and 34, corresponding to H-deficient HO_2 and H_2O_2 species, consistent with hydrogen loss. The background at mass 17, caused by fragmentation of water molecules in the mass spectrometer, did not allow us to make inferences about the ejection of OH radicals. We also found that sublimation from irradiated ice at 80 K is orders of magnitude larger than from unirradiated ice.

The values of Y at saturation are comparable to the probability⁸ of photoabsorption ($\sim 8 \times 10^{-3}$) in the first surface layer. This implies that photodesorption is very efficient, but not until the surface layers have been altered. Assuming the photoabsorption cross-section does not change significantly for

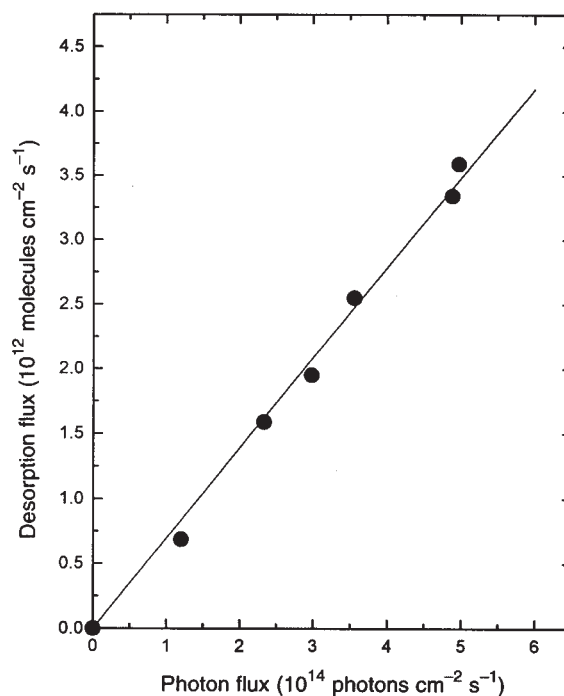


FIG. 1 Desorption flux from an ice film at temperature $T = 100$ K plotted against Lyman- α flux on the sample for an accumulated dose of $\sim 5 \times 10^{18}$ photons cm^{-2} . The linear behaviour shows the absence of thermal effects.

the altered material, each molecule will have absorbed a photon and reformed (on average) a number of times, before contributing to a desorption event. This indicates that ice is very stable against radiation even at the highest doses of deposited energy used (~ 480 eV per molecule averaged over the penetration depth).

Comparison of our results with existing estimates of Y for ice is difficult because neither the dose nor the temperature dependence were anticipated earlier. Our yields at saturation are much lower than the value of ~ 0.1 estimated recently by Hartquist and Williams⁶, but are much higher than earlier estimates of 4×10^{-5} by Draine and Salpeter⁵. The measured yield is close to Harrison and Schoen's² estimate in the saturation region at low temperatures, an exceeds it by a factor of ~ 3 at 100 K.

In interstellar clouds, the expected fluxes of energetic ultraviolet photons are $F = 10^8 \text{ cm}^{-2} \text{ s}^{-1}$ for the background photon field expected in diffuse and translucent interstellar clouds or on the outskirts of dense clouds¹⁶. After a dose of 10^{18} photons cm^{-2} (Fig. 3) or a time of ~ 300 years, the desorption rate from an ice mantle is then $> 2 \times 10^5 \text{ molecules cm}^{-2} \text{ s}^{-1}$, several orders of magnitude larger than the gas adsorption rate in a diffuse cloud, implying rapid removal of a pre-existing ice mantle. This explains the puzzling absence of the characteristic infrared absorption at a wavelength of $3.1 \mu\text{m}$ (the 'ice signature') in these diffuse interstellar environments (visual extinction $< \sim 3.3$)¹⁷.

Our measurements also imply that photodesorption is an important mechanism at the outskirts of dense molecular clouds, where the hydrogen density is $n_{\text{H}} = 10^3\text{--}10^5 \text{ cm}^{-3}$, or in clumpy molecular clouds where the ultraviolet radiation can penetrate. At the core of dark clouds, the interstellar ultraviolet flux is attenuated, but a flux of a few thousand photons per cm^2 per s is believed to exist due to excitation of background H_2 by cosmic rays¹⁸ and shocked stellar winds¹⁹. At these low levels, the photo-

desorption flux is $\sim 10^6\text{--}10^7 \text{ molecules cm}^{-2} \text{ yr}^{-1}$, several orders of magnitude smaller than the accretion rate ($10^{11}\text{--}10^{14} \text{ molecules cm}^{-2} \text{ yr}^{-1}$). Therefore in the absence of additional processes, photodesorption cannot remove much of a water-ice grain mantle in a dense region of a molecular cloud, consistent with recent observations²⁰.

We now examine implications for circumsolar ice grains. Assuming that yields simply scale in the ultraviolet with the photoabsorption cross-section, we can scale estimates² for photodesorption of ice in the outer Solar System using our Lyman- α data. At saturation, the calculated photodesorption rate, using the data in Fig. 2, exceeds the solar-wind sputtering calculated earlier²¹ and, beyond 5 AU, is larger than sublimation. Near Uranus (~ 70 K) the photodesorption rate after ~ 40 years of exposure of an undisrupted ice surface is $\sim 50\%$ larger than estimated earlier²², whereas near Saturn (~ 100 K) it is about three times larger than the estimate² after only ~ 10 years of exposure. This is close agreement, given the uncertainties in extrapolating laboratory data to astronomical conditions. For ice embedded in planetary magnetospheres, like Saturn's E-ring particles^{23,24} and the icy satellites²⁵, erosion by photons, although not negligible, is less important than by plasma ions²⁵.

These desorption rates are a lower bound because other processes can contribute synergistically to desorption. Sputtering of ice by magnetospheric ions may be enhanced because of precursor radicals produced by ultraviolet irradiation, possibly explaining the deficit between the observation of OH near Tethys²⁶ and sputtering calculations²⁷. In addition, enhanced sublimation of photolysed ice may be triggered by a transient temperature increase produced by a grain-grain collision²⁸ or a cosmic ray²⁹. This is supported in part by our observation of enhanced sublimation mentioned above; however, we did not find 'explosive' desorption in water ice because the back-reaction to reform water is apparently very efficient.

Although our data showed that the saturation Y was independent of photon flux in the range $10^{14}\text{--}10^{15} \text{ cm}^{-2} \text{ s}^{-1}$, this does

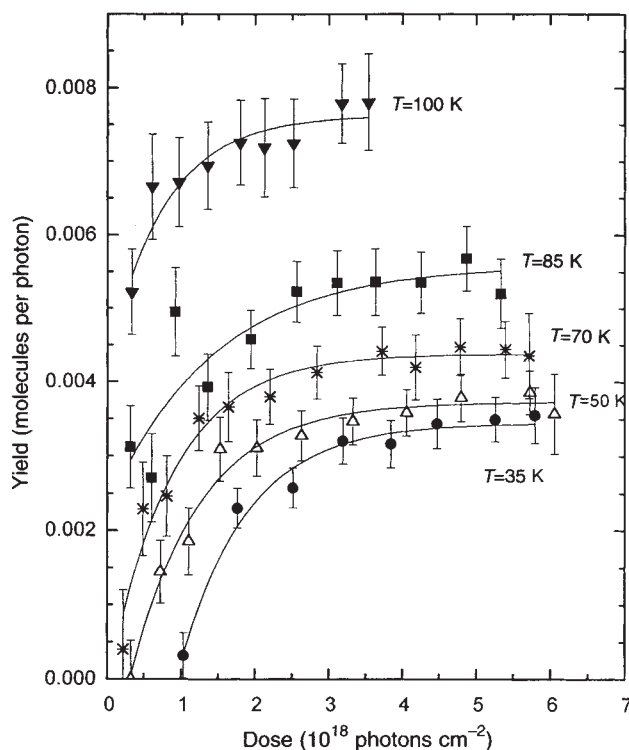


FIG. 2 Photodesorption yield of water ice as a function of Lyman- α photon dose, for different temperatures T . The growth and irradiation temperatures were the same. Photon fluxes during measurements were in the range $(0.5\text{--}5) \times 10^{14} \text{ photons cm}^{-2} \text{ s}^{-1}$.

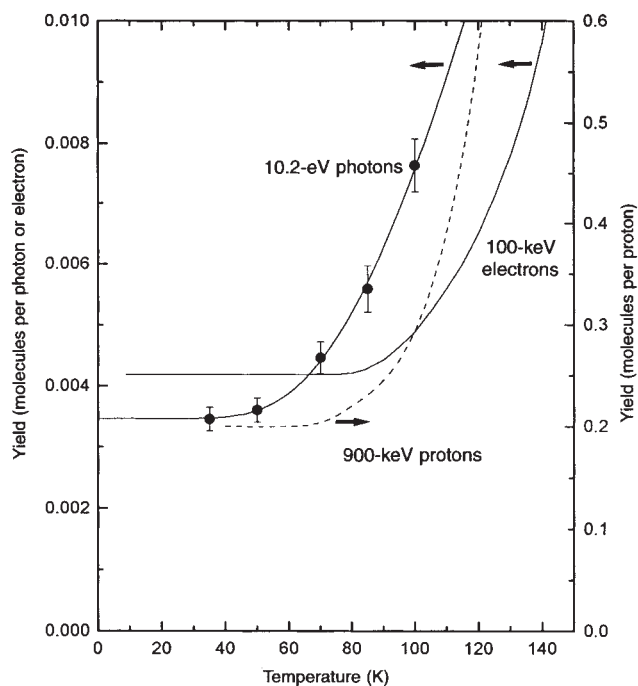


FIG. 3 Photodesorption yield of water ice as a function of temperature for Lyman- α photons. Also shown are desorption yields for fast protons^{11,12} and electrons¹⁴. The line through the photodesorption data is a fit to $Y(T) = Y_0 + Y_1 e^{-E/kT}$. Fitted values are $Y_0 = 0.035 \pm 0.002$, $Y_1 = 0.13 \pm 0.10$, $E = (29 \pm 6) \times 10^{-3} \text{ eV}$.

not necessarily imply that situation will hold in photon fluxes of 10^4 – 10^8 cm $^{-2}$ s $^{-1}$ expected in the astrophysical environments of icy grains. It is possible that, in this case, the radicals may recombine in the time of 1 – 10^4 years between photon impacts in a region of molecular dimensions (10^{-15} cm 2). To elucidate this effect, which would result in smaller Y values than those reported here, detailed modelling based on the photon flux and the measured activation energy will be performed. $\square$

Received 5 July; accepted 5 December 1994.

1. Sternberg, A., Dalgarno, A. & Lepp, S. *Astr. J.* **320**, 676–682 (1987).
2. Harrison, H. & Schoen, R. I. *Science* **157**, 1175–1176 (1967).
3. Carlson, R. W. *Nature* **283**, 461 (1980).
4. Combi, M. R. *Icarus* **71**, 178–191 (1987).
5. Draine, B. T. & Salpeter, E. E. *Astr. J.* **231**, 438–455 (1979).
6. Hartquist, T. W. & Williams, D. A. *Mon. Not. R. astr. Soc.* **247**, 343–348 (1990).
7. Sack, N. J. & Baragiola, R. A. *Phys. Rev.* **B48**, 9973–9978 (1993).
8. Warren, S. G. *Appl. Opt.* **23**, 1206–1225 (1984).
9. Nishi, N., Shinohara, H. & Okuyama, T. *J. chem. Phys.* **80**, 3898–3910 (1984).
10. Watson, W. D. & Salpeter, E. E. *Astr. J.* **174**, 321–340 (1972).

11. Brown, W. L., Lanzerotti, L. J., Poate, J. M. & Augustyniak, W. M. *Phys. Rev. Lett.* **40**, 1027–1030 (1978).
12. Brown, W. L., Augustyniak, W. M., Lanzerotti, L. J., Johnson, R. E. & Evatt, R. *Phys. Rev. Lett.* **45**, 1632–1635 (1980).
13. Taub, I. A. & Eiben, K. J. *chem. Phys.* **49**, 2499–2513 (1968).
14. Heide, H.-G. *Ultramicroscopy* **14**, 271–278 (1984).
15. Flournoy, M., Baum, L. H. & Siegel, S. J. *chem. Phys.* **36**, 2229–2230 (1962).
16. Mathis, J. S., Mezger, P. G. & Panagia, N. *Astr. Astr.* **128**, 212–229 (1983).
17. Whittet, D. C. B. *Dust in the Galactic Environment* (Inst of Physics, Bristol, 1992).
18. Prasad, S. S. & Tarafdar, S. P. *Astr. J.* **267**, 603–609 (1983).
19. Norman, C. & Silk, J. *Astr. J.* **238**, 158–174 (1980).
20. Smith, R. G., Sellgren, K. & Brooke, T. Y. *Mon. Not. R. astr. Soc.* **263**, 749–766 (1993).
21. Brown, W. L., Lanzerotti, L. J. & Johnson, R. E. *Science* **218**, 525–531 (1982).
22. Eviatar, A. & Richardson, J. *Astr. J.* **300**, L99–L102 (1986).
23. Morfill, G. E., Havnes, O. & Goertz, C. K. *J. geophys. Res.* **98**, 11285–11297 (1993).
24. Hamilton, D. & Burns, J. *Science* **264**, 550–553 (1994).
25. Johnson, R. E. *Energetic Particle Interaction with Atmospheres and Surfaces* (Springer, Berlin, 1990).
26. Shemansky, D. E., Matheson, P., Hall, D. T., Hu, H.-Y. & Tripp, T. M. *Nature* **363**, 329–331 (1993).
27. Johnson, R. E., Grosjean, D. E., Jurac, S. & Baragiola, R. A. *EOS* **74**, 572–573 (1993).
28. d'Hendecourt, L. B., Allamandola, L. J., Baas, F. & Greenberg, J. M. *Astr. Astrophys.* **109**, L12–L14 (1982).
29. Leger, A., Jura, M. & Omont, A. *Astr. Astrophys.* **144**, 147–160 (1985).

ACKNOWLEDGEMENTS. We thank I. Harrison and B. E. Turner for discussions, and B. Draine for valuable criticism. This work was supported by the US NSF, a NASA graduate Research Fellowship (M.S.W.) and the Italian Space Agency (G.A.B.).

A structural phase transition induced by an external magnetic field

A. Asamitsu*, Y. Moritomo*, Y. Tomioka*,
T. Arima† & Y. Tokura*†

* Joint Research Center for Atom Technology (JRCAT), Tsukuba, Ibaraki 305, Japan

† Department of Physics, University of Tokyo, Tokyo 113, Japan

A VAST number of compounds are known that exhibit structure transformations in response to changes in temperature, pressure and/or composition. One such example is the family of perovskites, $\text{La}_{1-x}\text{Sr}_x\text{MnO}_3$; for a limited range of compositions (x), they undergo a structural phase transition from an orthorhombic to a rhombohedral form with increasing temperature^{1,2}. These compounds are also ferromagnetic, a property that arises from coupling between the charge carriers and localized spin moments of the manganese ions^{3–7}. Here we show that, through careful tuning of the composition, the local spin moments and the charge carriers can in turn be coupled strongly to changes in the structure. For $x=0.170$, the crystal structure of the compound can be switched—reversibly or irreversibly, depending on the temperature—by application of an external magnetic field.

The ferromagnetic ground state of metallic $\text{La}_{1-x}\text{Sr}_x\text{MnO}_3$ (refs 3, 4) is caused by the so-called “double-exchange interaction” between the Mn^{3+} and Mn^{4+} ions^{5–7}. The parent anti-ferromagnetic insulator LaMnO_3 contains Mn^{3+} ions with $t_{2g}e_g^1$ (spin quantum number $S=2$) configuration. The t_{2g} electrons are localized and so can be viewed as local spins with $S=3/2$. In contrast, the e_g electrons are strongly hybridized with the oxygen $2p$ states and change from a localized to itinerant character with oxidation of the MnO_3 lattice⁸. An important feature is that strong intra-atomic ferromagnetic (Hund) coupling acts between the t_{2g} spins and e_g electrons, which gives rise to a parallel alignment of their spins. Chemical substitution of La^{3+} by Sr^{2+} introduces holes into the e_g orbitals, that are mobile and mediate an interatomic ferromagnetic interaction between the Mn atoms; the e_g electrons gain kinetic energy through this double-exchange interaction. This scheme was proposed by de Gennes⁷ as a microscopic origin of the ferromagnetic metallic states in the hole-doped LaMnO_3 perovskites. Very recently,

such a strongly spin-charge coupled state has been extensively revisited in light of the observation of a colossal negative magnetoresistive effect^{1,9–13}. Near the ferromagnetic ordering temperature (T_c), an external magnetic field can align more or less disordered local spins ferromagnetically. The forcedly spin-polarized conduction electrons (holes) then suffer less from the scattering by local spins and become more itinerant.

In addition to strong spin-charge coupling, changes in the crystal structure can also couple strongly to the electronic states. The lattice of $\text{La}_{1-x}\text{Sr}_x\text{MnO}_3$ undergoes a structural phase transition between the orthorhombic form (low-temperature phase, $Pbnm$) and the rhombohedral form (high-temperature phase, $R\bar{3}c$) with a change of the composition x and/or the temperature^{1,2}. We show in Fig. 1 the schematic crystal structure

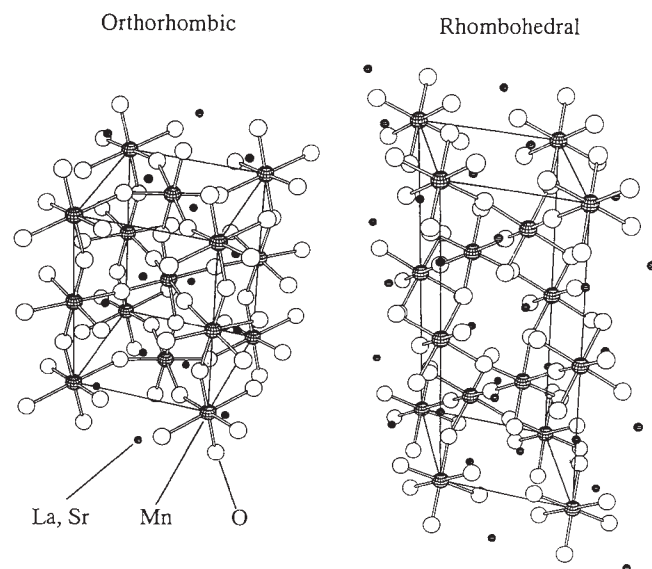


FIG. 1 Schematic crystal structures of $\text{La}_{1-x}\text{Sr}_x\text{MnO}_3$ ($x=0.170$) in the orthorhombic phase ($Pbnm$) and the rhombohedral phase ($R\bar{3}c$). The lattice parameters for the rhombohedral phase (measured at 290 K) are $a=5.475$ (1) Å and $\alpha=60.997^\circ$, giving a volume of 59.290 Å 3 per Mn atom. For the orthorhombic phase (measured at 280 K), the lattice parameters are $a=5.502$ (1) Å, $b=5.547$ (1) Å and $c=7.790$ (1) Å. The corresponding volume is 59.445 Å 3 per Mn atom.

in the orthorhombic and rhombohedral phases. A thermally induced structural change is observed in the composition region of $0.15 < x < 0.20$, and takes place at room temperature for values of x around 0.170. The main cause of such a lattice transformation is a change of the respective ionic sizes of the constituent ions, or the tolerance factor of the perovskite-like structure¹⁴. The transfer interaction of the e_g conduction electrons (holes) is expected to be larger in the rhombohedral lattice than in the orthorhombic one, as a larger distortion of the Mn–O–Mn bond angle (that is, deviation from 180°) in the latter lattice tends to decrease the transfer interaction. This fact is well known for the e_g electron-based conductors, it shows up most dramatically in the metal–insulator transition of the LnNiO_3 perovskites¹⁵ (where Ln is a rare earth). Thus not only the local spins, but also the lattice degree of freedom is expected to couple with the kinetic energy of the e_g electron (hole) carriers.

On the basis of the above scheme, we propose that the structural phase transition at a fixed temperature can be triggered by application or removal of a magnetic field in the following manner. The applied magnetic field directly acts on the t_{2g} local spin system. Then, as observed in the giant magnetoresistance effect of this system¹, the transfer energy of the e_g electron (hole) is increased, so that the lattice structure may show a field-induced transformation from the orthorhombic to rhombohedral form. Magnetic-field-induced switching of the crystal structure should be most effective if the structural transition temperature (T_s) is close to the ferromagnetic ordering temperature (T_c). This could be achieved by careful fine tuning of the composition x .

Crystals of $\text{La}_{1-x}\text{Sr}_x\text{MnO}_3$ were grown by the floating-zone method. A mixture of La_2O_3 , SrCO_3 and Mn_2O_3 with a prescribed ratio was prepared as a starting material. After the sintering process, crystals of $\text{La}_{1-x}\text{Sr}_x\text{MnO}_3$ were melt-grown by a floating-zone furnace in an appropriate flowing gas, the details of which will be described elsewhere. Fine control of the composition in steps of $x=0.005$ was required to optimize the field-switching characteristics of the crystal structure, as T_s decreases sharply with increasing x ($T_s=380$ K at $x=0.150$; $T_s=190$ K at $x=0.175$), whereas T_c rises from 238 to 283 K (ref. 1).

To characterize the crystal, we measured X-ray powder diffraction patterns and performed electron-probe microanal-

ysis. The analyses indicated that the crystals were single-phase with a composition nearly identical to the prescribed one. The temperature dependence of the X-ray powder diffraction pattern was also measured to confirm the structural phase transition with change of temperature. The lattice parameters for the rhombohedral and orthorhombic phases for the $x=0.170$ crystal, measured at 290 and 280 K respectively, are given in Fig. 1 legend. The structural phase transition was further pursued by measuring the change of the orthorhombic (113) peak intensity with temperature. The temperature dependence of the peak intensity in zero field showed hysteresis, and the values of T_s in both cooling and warming runs were in agreement with those determined by resistivity and lattice striction behaviour of this crystal (Fig. 2). At zero field, $T_c=264$ K, as determined by measurements of a.c. susceptibility. The decrease in resistivity below T_c arises from the phase transition from a paramagnetic insulator to a ferromagnetic conductor. The abrupt increase of resistivity and lattice striction around $T=280$ K in the cooling run corresponds to the structural phase transition from the rhombohedral to orthorhombic form, and the decrease around $T=290$ K in the warming run corresponds to the reverse transformation. The hysteresis observed in both resistivity and lattice striction (hatched areas in Fig. 2), reflect the first-order nature of this transition. Although it is difficult to determine the exact direction of lattice striction using a uniaxial strain gauge, the observed value of $\delta L/L \approx (3-4) \times 10^{-4}$, where L is length, at the transition is compatible with the volume change per Mn-site, $\delta V/V \approx 26 \times 10^{-4}$, in terms of the relation, $3\delta L/L = \delta V/V$.

With application of magnetic field, the resistivity drops drastically, showing the giant magnetoresistance effect in this crystal. For example, the negative magnetoresistance value, defined as $[\rho(0) - \rho(H)]/\rho(0)$ where $\rho(0)$ is zero-field resistivity and $\rho(H)$ is magnetoresistivity, reaches values as large as 0.91 for $\mu_0 H = 7$ T at 273 K. The structural phase-transition temperature decreases continuously with increasing field and the hysteresis loop becomes wider. At high fields ($\mu_0 H = 7$ T), the loop occurs below T_c .

Figure 3 shows the structural phase diagram of $\text{La}_{1-x}\text{Sr}_x\text{MnO}_3$ ($x=0.170$) determined in the plane of temperature–magnetic field. The value of T_c at zero field is also indicated. The region surrounded by the two phase-boundary lines is the bistable region where both the rhombohedral and orthorhombic forms can occur because of the hysteretic nature of the structural transition, which depends on whether the temperature is

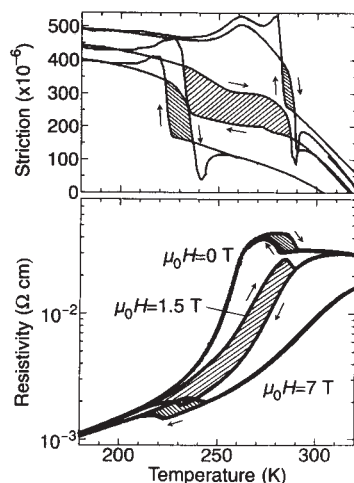


FIG. 2 Temperature dependence of resistivity (ρ , lower panel) and lattice striction ($\delta L/L$, upper panel) in $\text{La}_{1-x}\text{Sr}_x\text{MnO}_3$ ($x=0.170$) at various magnetic fields. The hatched regions represent the hysteresis loop of the structural phase transition between the rhombohedral form (high-temperature phase) and orthorhombic one (low-temperature phase). The arrows indicate the direction of temperature sweep. The resistivity and the lattice striction were monitored simultaneously.

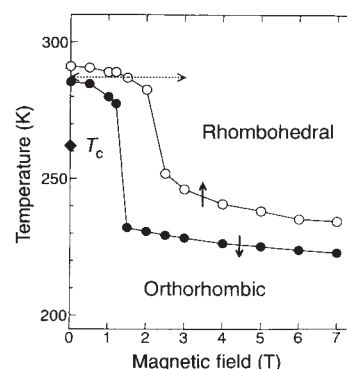


FIG. 3 Structural phase diagram of $\text{La}_{1-x}\text{Sr}_x\text{MnO}_3$ ($x=0.170$) in the plane of temperature–magnetic field. The structural phase-transition temperatures from the rhombohedral ($R3c$) to the orthorhombic ($Pbnm$) phase in the cooling run are shown by closed circles, and open circles indicate the reverse transformation in the warming run; the arrows give the direction of the temperature sweep. A broken line at $T=287.0$ K corresponds to the magnetoresistance measurements in Fig. 4 as a demonstration of irreversible switching of the crystal structure with a magnetic field.

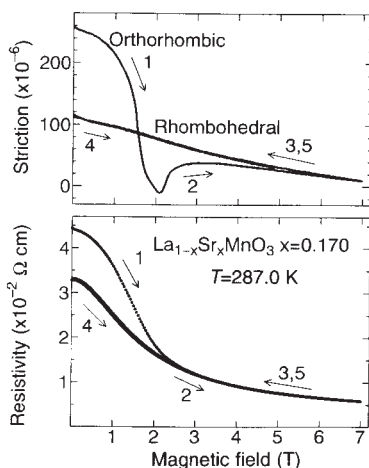


FIG. 4 Resistivity and lattice striction of $\text{La}_{1-x}\text{Sr}_x\text{MnO}_3$ ($x=0.170$) at 287.0 K as a function of magnetic field. The arrows with numbers indicate the sequence of field variation. The behaviour shows irreversible structural switching by external magnetic field from the orthorhombic to rhombohedral form, as well as the switching of the resistive state in this system.

increased or decreased. As the magnetic field is increased, T_s abruptly decreases across T_c and then becomes less field-dependent.

When the temperature is set within the hysteresis loop, it is possible to irreversibly switch the crystal structure by an external magnetic field. To demonstrate structural switching, we show in Fig. 4 the magnetic-field dependence of resistivity and lattice striction at 287.0 K. After cooling in zero field, the sample temperature is raised to 287.0 K in the hysteretic region so that the crystal is in the orthorhombic phase. The magnetic field is then gradually increased at 287.0 K, as indicated by a horizontal broken arrow in Fig. 3. The lattice contracts by $\sim 0.02\%$ around $\mu_0 H = 1.5$ T, showing the structural phase transition from the orthorhombic to the rhombohedral form. An overshooting behaviour in the lattice striction, perhaps originating from the first-order nature of this field-induced phase transition, is gradually recovered with increasing field up to $\mu_0 H = 7$ T. The resistivity also drops rapidly in accord with the structural switching, though the normal giant magnetoresistance effect also contributes to the decrease in resistivity. In the field-decreasing process, the resistivity and the lattice striction trace distinct curves as the crystal remains in the rhombohedral form. The second field-increasing run makes no further change in the both curves, showing that the irreversible switching of the crystal structure is completed. The resultant differences at zero field in ρ and $\delta L/L$ are considered to be caused by the structural transformation. Thus the present field-switching of the crystal structure also causes an irreversible switching of the resistive state, which may be viewed as a new magnetoresistive effect. It should be also noted that the $\delta L/L$ in the second run indicates large magnetostriction though the crystal remains in the rhombohedral phase. This may suggest the possibility of control of lattice parameters ("giant magnetostriction") in this system. $\square$

Received 9 November; accepted 19 December 1994.

1. Tokura, Y. et al. *J. phys. Soc. Japan* **63**, 3931–3935 (1994).
2. Wold, A. & Arrott, R. J. *J. Phys. Chem. Solids* **9**, 176–180 (1959).
3. Jonker, G. H. & Van Santen, J. H. *Physica* **16**, 337–349 (1950).
4. Jonker, G. H. *Physica* **22**, 707–722 (1956).
5. Zener, C. *Phys. Rev.* **82**, 403–405 (1951).
6. Anderson, P. W. & Hasegawa, H. *Phys. Rev.* **100**, 675–681 (1955).
7. de Gennes, P.-G. *Phys. Rev.* **118**, 141–154 (1960).
8. Goodenough, J. B. *Prog. Solid State Chem.* **5**, 145 (1971).
9. Kusters, R. M., Singleton, D. A., McGreevy, R. & Hayes, W. *Physica B* **155**, 362–365 (1989).
10. Chahara, K., Ohno, T., Kasai, M. & Kozono, Y. *Appl. Phys. Lett.* **63**, 1990–1992 (1993).

11. von Helmont, R., Wecker, J., Holzapfel, B., Schultz, M. & Samwer, K. *Phys. Rev. Lett.* **71**, 2331–2333 (1993).
12. Jin, S. et al. *Science* **264**, 413–415 (1994).
13. Furukawa, N. *J. phys. Soc. Japan* **63**, 3214–3217 (1994).
14. Goodenough, J. B. & Longon, J. M. *Landolt-Bornstein Tabellen* Vol. III/4a (Springer, Berlin, 1970).
15. Torrance, J. B., Lacorre, P. & Nazzari, A. I. *Phys. Rev. B* **45**, 8209–8212 (1992).

ACKNOWLEDGEMENTS. We thank N. Hamada for permission to use the drawings of crystal structure. This work was supported by the New Energy and Industrial Technology Organization (NEDO) of Japan.

Chlorofluorocarbon evidence for recent ventilation of the deep Bering Sea

Mark J. Warner & Gunnar I. Roden

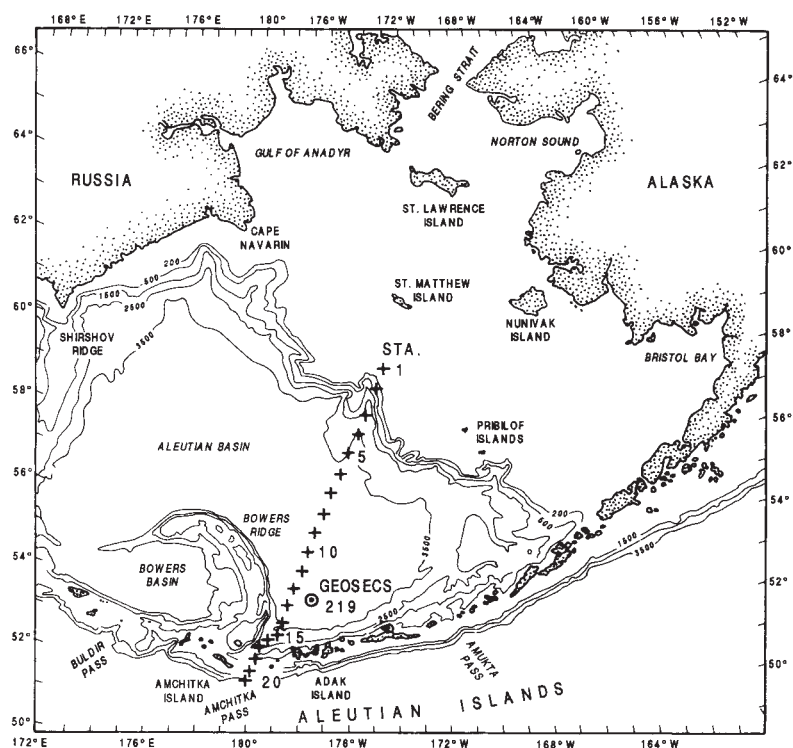
School of Oceanography, WB-10, University of Washington, Seattle, Washington 98195, USA

THE pattern of the global thermohaline circulation of today's oceans is controlled by deep-water formation at the northern and southern limits of the Atlantic. The apparent lack of deep-water formation in the North Pacific, on the other hand, suggests that this ocean plays only a minor role in the global circulation, but it is unclear whether this was also the case during periods of glaciation—there is conflicting evidence in the sedimentary record for the existence of a deep-water source in the North Pacific during the Last Glacial Maximum^{1–4}. Here we report the detection of anthropogenic chlorofluorocarbons in the bottom waters of the Aleutian basin in the eastern Bering Sea, which suggests that a small amount of bottom water has formed in this region during the past 40 years. Although the small volumes of water involved are unlikely to play a significant role in determining present-day global circulation patterns, the results lend credence to the possibility that this sea was an important source of deep water for the northwestern Pacific Ocean during the Last Glacial Maximum.

These measurements were made during July 1993 as part of the World Ocean Circulation Experiment (WOCE) hydrographic programme section P14N aboard the RV *Thomas G. Thompson*. Twenty hydrographic stations were occupied in the Bering Sea (Fig. 1) with samples collected throughout the water column from the surface to within 10 m of the bottom using a 36-bottle rosette instrumented with a Neil Brown (Cataumet, MA, USA) conductivity-temperature-pressure probe (CTD) and an oxygen sensor. Every water sample was analysed for dissolved oxygen, salinity and nutrients (nitrate, nitrite, phosphate and silicate). The chlorofluorocarbons CFC-11 and CFC-12 were analysed in $\sim 60\%$ of the water samples using established techniques⁵. The data are still preliminary, but large corrections in the final data set are not expected. The CFCs have been used as tracers of ocean circulation and ventilation in many studies over the past decade^{6–8} mainly because of their stability in the water column and their fairly well known source functions. In this study, the large dynamic range of CFC concentrations, from the surface-water CFC-11 values of >5 pmol kg^{-1} to the detection limit of 0.010 pmol kg^{-1} , enables us to identify the possible source of bottom water to the Aleutian basin.

The circulation and source of deep water in the Bering Sea are poorly understood, mainly due to a lack of high-quality deep hydrographic data. Kamchatka Strait is the only location where waters deeper than 2,000 m can exchange between the North Pacific and the Bering Sea. The waters of the Aleutian basin at depths greater than 3,500 m are further isolated from North Pacific influence by the Shirshov and Bowers ridges (Fig. 1). The deep waters of the Bering Sea are believed to be the oldest of the world's oceans and they are noted for extremely high silicate concentrations⁹ which are the result of high productivity

FIG. 1 Bathymetry of the Bering Sea¹⁷ showing positions of stations (crosses) occupied on a segment of World Ocean Circulation Experiment hydrographic program P14N aboard the RV *Thomas G. Thompson*. These measurements were made during 5–15 July 1993. GEOSECS station 219 was occupied on 7 October 1973.



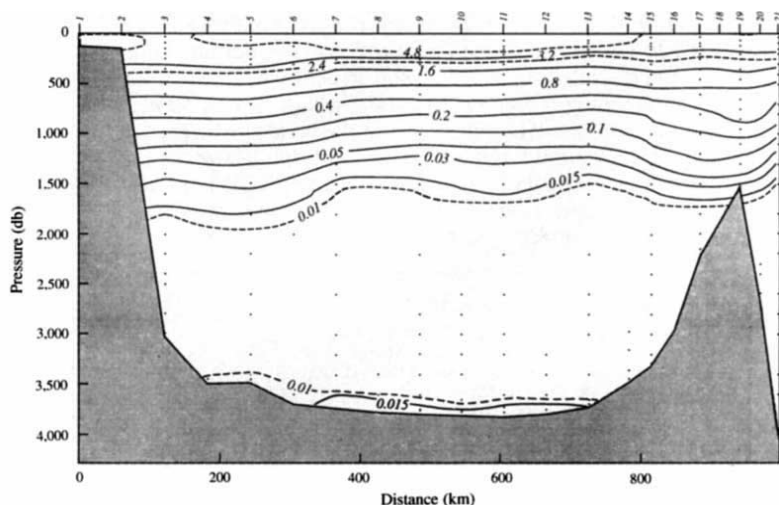
and dissolution from the sediments¹⁰. Other properties are also affected by the degradation of organic matter in the basin. The Bering Sea waters are lower in oxygen and higher in dissolved nutrients than the deep North Pacific at the same densities. Previous studies have hypothesized that Pacific deep water flows slowly into the basin^{10–12} where it mixes and return to the Pacific at an intermediate depth. The inflow appears to be small enough that geothermal warming could be used to trace the flow of bottom water (T. E. Whitledge, personal communication). Previous high-quality CTD measurements have revealed a 100-m-thick layer of water at the bottom which differs in potential temperature and dissolved oxygen from the overlying waters at GEOSECS station 219¹³ and other stations in the Aleutian basin (T. E. Whitledge, personal communication). Records of potential temperature and oxygen (from CTD measurements) change abruptly ~120 m above the sea floor and then remain nearly constant in the lower 60 m. Dissolved silicate in this layer was found to decrease with depth, implying that this water is younger than the overlying water. The presence of this bottom layer has

been interpreted as resulting from replenishment with waters from the North Pacific¹³ or a combination of inflow and flux from the sediments¹².

During this WOCE expedition, this bottom water was found at all stations deeper than 3,300 m (stations 4–14, Fig. 1). Concentrations of CFC-11 greater than the detection level of 0.010 pmol kg⁻¹ were found in the bottom water beginning at station 5 with the highest concentrations at station 11 (Fig. 2). The precision of these measurements is ± 0.005 pmol kg⁻¹ and a sampling blank of 0.006 pmol kg⁻¹ has been subtracted to correct for contamination. Because the CFC-11/CFC-12 ratio of sea water in equilibrium with the atmosphere is ~2, the CFC-12 concentrations are generally below the detection limit.

Profiles of CTD potential temperature in the bottom layer on this section are similar to the GEOSECS station 219 data¹³, but the silicate concentrations only exhibit a maximum above the bottom at a few stations (Fig. 3). Higher CFC-11 concentrations within this bottom water are correlated with higher dissolved oxygen, lower silicate, lower potential temperature and greater

FIG. 2 Vertical section of CFC-11 concentration (contours, in units of pmol kg⁻¹) along the section in Fig. 1. Station numbers are given along the upper horizontal axis, and the CFCs were analysed at the marked pressures (dots). Note that the solid contour intervals reflect a doubling of concentration. The dashed line at 0.010 pmol kg⁻¹ is the detection limit. North is to the left. Vertical exaggeration is 143:1. db, decibars.



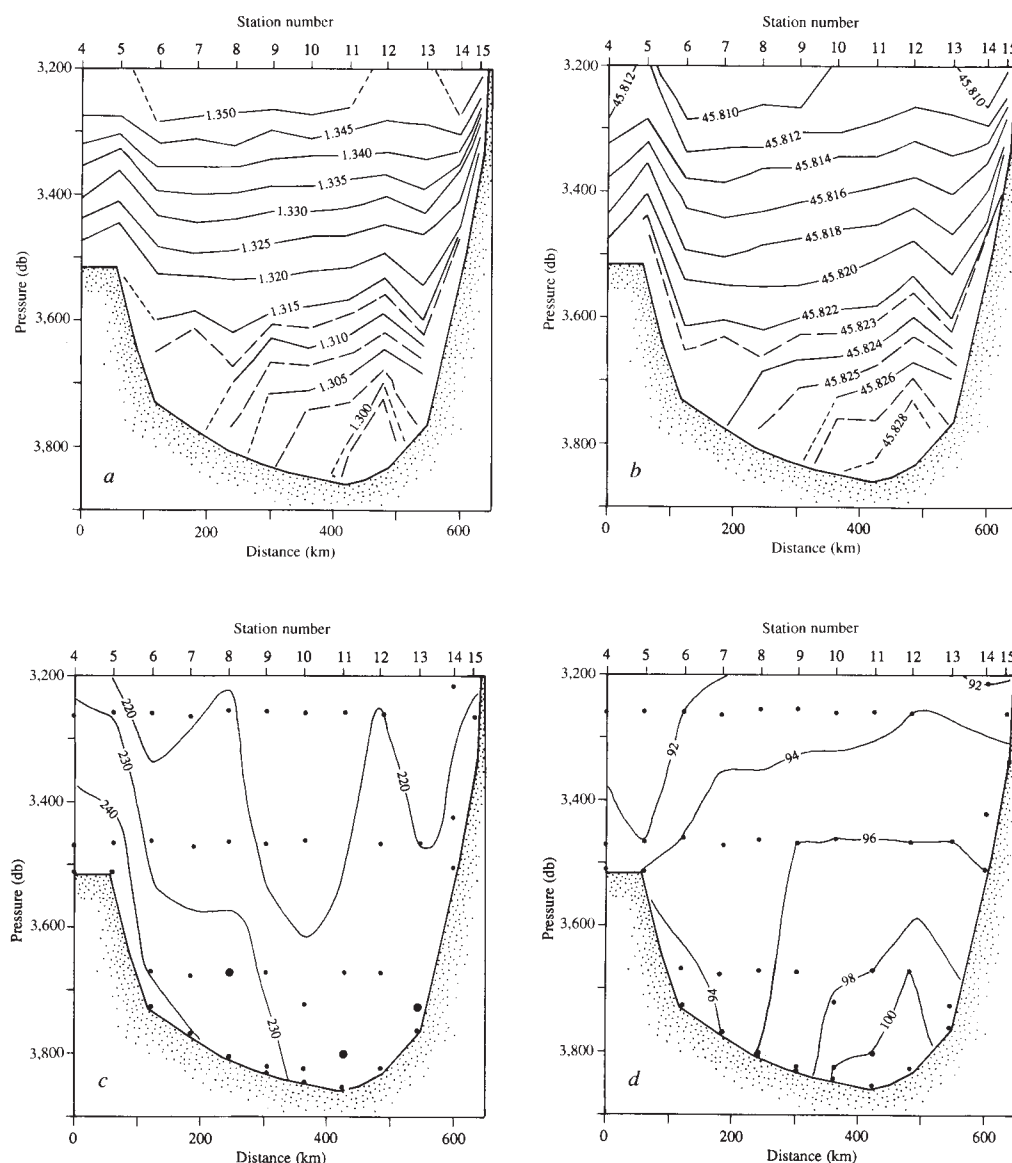


FIG. 3 Vertical sections of: *a*, CTD potential temperature ($^{\circ}\text{C}$); *b*, CTD potential density relative to 4,000 db; *c*, dissolved silicate ($\mu\text{mol kg}^{-1}$); *d*, dissolved oxygen ($\mu\text{mol kg}^{-1}$) at stations with bottom pressure greater than 3,200 db (that is, stations 4–14 in the previous figures). The dots in *c* and *d* indicate sample locations, the larger dots in the former indicating the samples with the highest concentrations at each station. North is to the left and the vertical exaggeration is 1,000:1. Note that the distance scale begins at station 4, whereas it begins at station 1 in Fig. 2.

density along the sea floor (Fig. 3). This water is also slightly more saline, with a maximum salinity of 34.670 (salinity based on the Practical Salinity Scale; ref. 18) at station 12. Except for the CFC-11 concentration, all the other parameters are consistent with the previous conclusions on the origin of this bottom water. However, as there are no CFCs at this density or depth in the North Pacific, the CFC-11 signal must be derived from within the Bering Sea.

We hypothesize that the bottom water is derived from the formation of a small amount of deep water in the Bering Sea. The major constraint on open-ocean deep convection in this region is the contrast between the low surface salinities and the deep salinities¹⁴. Shelf processes are a more probable mechanism for bottom-water formation. Salinities as high as 35 have been reported during polynyas on the continental shelf in the northern Bering Sea¹⁵. Although only a relatively small amount of dense water appears to be formed by a polynya, this water is dense enough to sink to the sea floor. Submarine canyons provide a conduit for this dense water to reach the deep Bering Sea without a great deal of mixing along its path.

Other possible sources for the CFC-11 signal can be eliminated using the available data. Adsorption and desorption of CFC-11 on sinking particles would affect the CFC-11/CFC-12 ratio higher in the water column, and would be negatively correlated with dissolved oxygen in the bottom water. Correlations of CFC-

11 concentrations with other parameters removes the possibility of a sunken refrigeration unit in the Aleutian basin, but they cannot eliminate the possible advection of CFC-11 into the Aleutian basin from a submerged source in one of the two basins between the North Pacific and this basin. A sunken vessel or discarded refrigerator would contain only one CFC and would therefore result in unusual CFC-11/CFC-12 ratios. CFC-11 and CFC-12 concentrations are positively correlated with a realistic ratio (1.1); however, the very low levels of CFC-12 render this relationship inconclusive.

Although strong evidence for deep-water formation has been found, the amount of water formed does not appear to be very large. The rate of deep-water production can be estimated if we assume that the dense water formed during one winter, that it sank into only the Aleutian basin, and that it was diluted tenfold with existing bottom waters. The resulting value, $1 \text{ sverdrup} \equiv 10^6 \text{ m}^3 \text{ s}^{-1}$ is comparable to the amount of dense water delivered annually to the Arctic Ocean by coastal polynyas¹⁶. The CFCs can be used to constrain the timescale over which this bottom-water formation has occurred. If the bottom water is assumed to be unmixed and have formed during one winter, the CFC-11 concentrations give a maximum time since deep-water ventilation of 41 yr. A more likely scenario is that a small amount of newly formed bottom water mixes with both the water it sinks through and the bottom waters in the

basin. In the case where the CFC-11 signal in the bottom water is modelled as the result of a CFC-bearing component diluted 100:1 with CFC-free water, the date of the deep-water formation would be 20 years ago.

These are the first CFC measurements in the bottom water of the Aleutian basin, and therefore cannot be used to address the temporal variability in deep-water formation. More realistic interpretation of the CFC age will depend upon models of the formation process and subsequent mixing into the bottom waters. The limited amount of previous high-quality hydrographic data also hinders analysis of the temporal variability in deep-water production and eventual fate of these ventilated waters. □

Received 21 July; accepted 30 November 1994.

1. Keigwin, L. D. *Nature* **330**, 362–364 (1987).
2. Duplessy, J. C. et al. *Paleoceanography* **3**, 343–360 (1988).
3. Boyle, E. A. *Rev. Earth Planet. Sci.* **20**, 245–287 (1992).
4. Naidu, A. S., Shackleton, N. J., Hall, M. & Maslin, M. *EOS (suppl.)* **74**, 322 (1993).
5. Bullister, J. L. & Weiss, R. F. *Deep-Sea Res.* **35**, 839–853 (1988).
6. Weiss, R. F., Bullister, J. L., Gammon, R. H. & Warner, M. J. *Nature* **314**, 608–610 (1985).
7. Wallace, D. W. R. & Lazier, J. R. N. *Nature* **332**, 61–63 (1988).
8. Warner, M. J. & Weiss, R. F. *Deep-Sea Res.* **39**, 2053–2075 (1992).
9. Mantyla, A. W. & Reid, J. L. *Deep-Sea Res.* **30**, 805–833 (1983).
10. Tsunogai, S., Kusakabe, M., Iizumi, H., Koike, I. & Hattori, A. *Deep-Sea Res.* **26**, 641–659 (1979).
11. Reed, R. K., Khen, G. V., Staben, P. J. & Verkhunov, A. V. *Deep-Sea Res.* **40**, 2325–2334 (1993).
12. Toggweiler, J. R. thesis, Columbia Univ., New York (1983).
13. Park, P. K., Broecker, W. S., Takahashi, T. & Reeburgh, W. S. in *Bering Sea Oceanography: An Update 1972–1974* (eds Takenouti, Y. & Hood, D. W.) 207–244 (Univ. Alaska, Fairbanks, 1975).
14. Warren, B. A. *J. Mar. Res.* **41**, 327–347 (1983).
15. Schumacher, J. D., Aagaard, K., Pease, C. H. & Tripp, R. B. *J. Geophys. Res.* **88**, 2723–2732 (1983).
16. Cavalieri, D. J. & Martin, S. J. *Geophys. Res.* **99**, 18343–18362 (1994).
17. Sayles, M. A., Aagaard, K. & Coachman, L. K. *Oceanographic Atlas of the Bering Sea Basin* (Univ. Washington, Seattle, 1979).
18. *UNESCO Technical Paper in Marine Science*, No. 37 (1981).

ACKNOWLEDGEMENTS. We thank the Oceanographic Data Facility at Scripps Institution of Oceanography for CTD/hydrographic data collection and processing. This research was supported by the Physical Oceanography division of the National Science Foundation.

Importance of iron for plankton blooms and carbon dioxide drawdown in the Southern Ocean

Hein J. W. de Baar*, Jeroen T. M. de Jong*,
Dorothee C. E. Bakker*, Bettina M. L  scher*,
Cornelis Veth*, Uli Bathmann†
& Victor Smetacek†

* Netherlands Institute for Sea Research, PO Box 59,
1790 AB Den Burg, Texel, The Netherlands

† Alfred Wegener Institute for Polar and Marine Research,
27570 Bremerhaven, Germany

THE iron hypothesis^{1–3}—the suggestion that iron is a limiting nutrient for plankton productivity and consequent CO₂ drawdown—has been tested by small-scale experiments in incubation bottles in the subarctic Pacific^{2,4} and Southern^{5–7} Oceans, and by a recent large-scale experiment in the equatorial Pacific Ocean^{8,9}. Here we test the idea by looking at natural levels of productivity in regions of the Southern Ocean with differing iron abundance. In the southerly branch of the Antarctic circumpolar current (ACC), upwelling of deep waters supplies sufficient iron to the surface to sustain moderate primary production but not to permit blooms to develop. In contrast, within the fast-flowing, iron-rich jet of the polar front (PF), spring blooms produced phytoplankton biomass an order of magnitude greater than that in southern ACC waters, leading to

CO₂ undersaturation. The plankton-rich PF waters were sharply delineated from adjacent iron-poor waters, indicating that iron availability was the critical factor in allowing blooms to occur.

The first iron experiments in the Weddell-Scotia Sea^{5,6} and the Ross Sea⁷ have shown stimulation of growth by iron; however, other limitations (light and grazing^{5,6,10,11}) were also at play in these near-shore, neritic¹² waters where dissolved Fe concentrations^{4,13,20} commonly exceed ~1 nM, consistent with the supply¹⁴ of this element from shelf sediments. To test the limitation hypothesis in truly oceanic waters of the Antarctic circumpolar current, well away from land sources, iron and its interaction^{15,16} with biota were studied along the 6° W meridian during the 1992 austral spring European IGBP-JGOFS expedition¹⁷ aboard the RV *Polarstern* (Fig. 1). We measured dissolved Fe concentrations in 120 seawater samples collected at six standard depths in conjunction with measurements of chlorophyll *a* (chl *a*), the partial pressure of CO₂ (*P*_{CO₂}) in sea water and marine air, along with observations by colleagues of many other variables¹⁷.

Vertical depth profiles (Fig. 2) show that the southern ACC waters have subnanomolar Fe concentrations, sometimes as low as 0.17 nM. The deepest (300-m) sample commonly exhibits the highest concentration, approaching ~1 nM, which is consistent with deep-water concentrations (Fig. 3). In surface waters at the PF subnanomolar Fe is found, but also a very pronounced maximum of 2–4 nM in the 60–100 m depth range. The maximum coincides with distinct hydrographic extrema (not shown) defining the polar frontal jet.

In the polar front (47–50° S), section 5/6 (Fig. 4a) shows the highest dissolved Fe: 1–3.5 nM in the upper 100 m of the water column, a minimum of ~0.5 nM at ~200 m depth and values of 1–1.5 nM at 300 m depth, which is approaching the deep-water concentrations. The PF three weeks later (section 11 in Fig. 4d) shows lower values of 1–1.5 nM in the upper 100 m. This seasonal decrease of ~1 nM in the PF is consistent with similar decreases of ~30 µatm for *P*_{CO₂} at 12 m depth (Fig. 4c and f), and for the upper 100 m of the water column, apparent

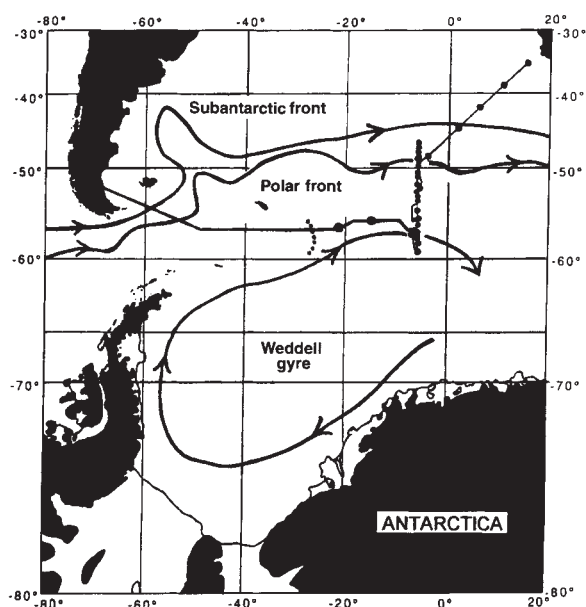


FIG. 1 Chart of the research area with initial west-east section 1 at ~56° S and several north-south sections (2–11) at ~6° W. Results of combined sections 5/6 (24 October–6 November 1992) and 11 (10–21 November) are reported, with some data from section 1. The polar front is a more rapidly eastward-flowing jet within the overall Antarctic circumpolar current (ACC), separating the wide southern branch ('ACC-proper') from the narrow northern ACC branch which was not part of this study.

TABLE 1 Upward transport versus aeolian input of iron

	ACC stations (52–58° S, 6° W)	Ocean station Papa (50° N, 145° W)
Advection flux*		
Velocity (cm s ⁻¹)	15 × 10 ⁻⁵ (ref. 32)	5 × 10 ⁻⁵ (ref. 33)
Maximum velocity (cm s ⁻¹)	(30 × 10 ⁻⁵) (ref. 34)	(10 × 10 ⁻⁵) (ref. 35)
Deep Fe concentration (mol m ⁻³)	1 × 10 ⁻⁶ (this work)	0.6 × 10 ⁻⁶ (ref. 35)
Advection flux (mol m ⁻² s ⁻¹)	1.5 × 10 ⁻¹²	0.3 × 10 ⁻¹²
Eddy-diffusive flux†		
K _z (best median values) (m ² s ⁻¹)	3 × 10 ⁻⁵ (refs 36, 37)	0.7 × 10 ⁻⁵ (refs 36, 38)
Gradient ∂[Fe]/∂z (mol m ⁻⁴)	6 × 10 ⁻⁹ (this work)	1.9 × 10 ⁻⁹ (ref. 35)
Eddy flux (mol m ⁻² s ⁻¹)	0.18 × 10 ⁻¹²	0.014 × 10 ⁻¹²
Total upward flux (mol m ⁻² s ⁻¹)	1.68 × 10 ⁻¹²	0.314 × 10 ⁻¹²
Sustained C production (mol m ⁻² s ⁻¹)	1.68 × 10 ⁻⁷	0.32 × 10 ⁻⁷
Sustained C production (mg C m ⁻² d ⁻¹)	170	32
Measured C production (mg C m ⁻² d ⁻¹)	80–300 (ref. 18)	
Annual total aerosol (mg m ⁻² yr ⁻¹)	3 (ref. 39)	30 (ref. 39)
10% dissolution (mol m ⁻² s ⁻¹)	0.18 × 10 ⁻¹²	1.8 × 10 ⁻¹²
Ice edge melting event MIZ‡	0.4 × 10 ⁻⁶	ND
Advection nitrate flux§	4.54	ND

Calculations of the annual upward transport of dissolved Fe in the ACC and the Gulf of Alaska. For the diffusivity K_z improved estimates^{36–38} were preferred over earlier approximations³⁵. Also given are order-of-magnitude assessments of annual inputs by aerosols, and by melting of 1-m-thick sea ice (Fe contents were found to be ~30 nM) into 75 m water column, as well as the upwelling supply of nitrate for its measured deep-water concentration of 35 μ M. ND, not determined.

* Defined as vertical velocity $\times$ Fe concentration.

† Defined as $K_z \partial[\text{Fe}]/\partial z$.

‡ Calculated from melting of 30 nM Fe in 1-m-thick sea-ice diluted into 75-m-deep mixed layer: $(30 \times 10^{-6} \text{ mol m}^{-3})/(75 \text{ m})$.

§ Calculated from $(15 \times 10^{-5} \text{ cm s}^{-1}) \times (35 \times 10^{-9} \text{ mol m}^{-3})$.

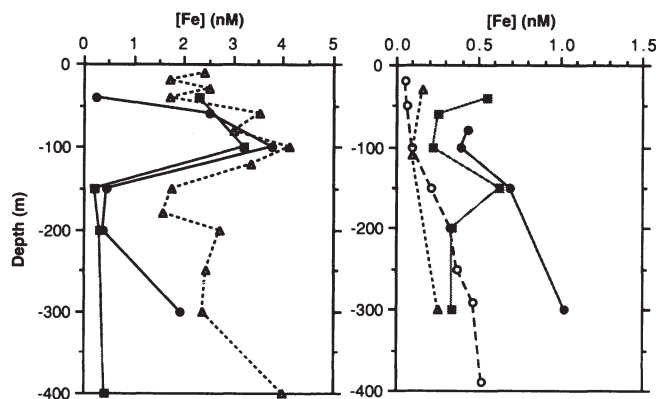


FIG. 2 Vertical profiles of dissolved Fe at 56° and 57° S (filled circles and squares, respectively) in the ACC (right) and 48° and 48.7° S (filled circles and squares, respectively) in the PF (left). Also shown on the right are the upper three (of five) values (triangles, short-dashed line) for Drake Passage²⁰ and seven values of the subarctic Pacific³⁵ (station Papa in Table 1; open circles, long-dashed line) and on the left, previous values for the Scotia Sea¹⁶ (open triangles, short-dashed line) as influenced by the Peninsula shelf source. Samples were collected at 40, 60, 100, 150, 200 and 300 m depth and analysed with methods described elsewhere⁴⁰. No samples shallower than 40 m as to avoid contamination from ships hull. The recent finding (ref. 41; E. L. Rue and K. W. Bruland, personal communication) in other seas of >99% organic complexation of iron and its effect on algal uptake was discussed elsewhere^{3,24}.

decreases¹⁷ of 10–15 μ M total CO₂, 1–2 μ M nitrate, ~0.1 μ M phosphate and ~10 μ M silicate (not shown). Meanwhile chl *a* levels had increased from 0.8–1.6 to 1.2–2.4 μ g l⁻¹ while the spring bloom evolved. Rates of primary productivity were observed¹⁸ to be ~1,200–3,000 mg C m⁻² d⁻¹. Nevertheless, macronutrients were still available in high concentrations (nitrate ~24 μ M, phosphate ~1.8 μ M, silicate ~21 μ M). Hence their complete utilization was hampered by other factors, such as light limitation due to intense storm mixing as well as grazing pressure. Other essential metals (Mn, Zn, Cu, Ni, Co) are deemed non-limiting: their cellular requirement is relatively less than for iron, and enrichment experiments¹⁶ showed little or no effect for these metals whereas surface-water concentrations of 'nutrient-type' metals Zn, Ni, Cu are high. But studies of the biological availability of these metals are needed to assess their true role.

The polar front region is the site of formation of Antarctic Intermediate Water (AAIW), and also the region of massive sedimentation of biogenic opal. Diatom blooms and sedimentation may be stimulated by the favourable ambient iron abundance. The iron also affects the relative uptake of nitrate and phosphate¹⁹. Hence the major-nutrients signature (silicate, nitrate, phosphate) of AAIW (flowing northwards into other ocean basins) is controlled by iron-regulation of diatom blooms.

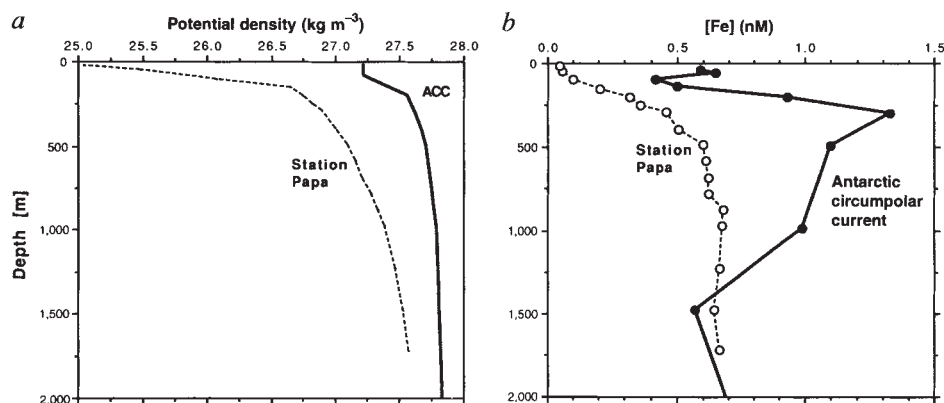
The polar front in both sections (5/6 and 11) exhibits ambient Fe: phosphate ratios of $(1.2\text{--}2.4) \times 10^{-3}$ which exceed the 'critical' 1:1,000 proportion⁵; whereas in the ACC, the ratios (Fe:P = ~0.2 $\times 10^{-3}$) are well below the 'critical' value, implying less favourable conditions for blooming.

The maxima of iron concentration in the PFZ are consistent with the high geostrophic velocity (~25 cm s⁻¹) of this zone compared with only ~5–10 cm s⁻¹ in the rest of the ACC section. The rapidly eastward flowing PFZ jet has still retained a significant signal of Fe input from shelf sources. Diagenetic input from shelf sediments is obviously significant within²⁰ and downstream¹⁴ of the Antarctic Peninsula archipelago and in the PF (Figs 1, 2). On passing through Drake Passage the polar frontal jet is known to shift to the northeast (Fig. 1) towards or across the Scotia ridge and in the 26–32° W region may have flowed over the northeast Georgian rise, picking up dissolved iron from sediment sources. When subsequently travelling another 20° eastward towards the 6° W meridian, the simultaneous removal of dissolved Fe from surface waters is thought to occur with settling biogenic particles. Iron is therefore expected to decrease eastwards, as seen at two stations in an east-west section (Fig. 1), where dissolved Fe at 50 m depth decreased from 1.3 nM (57° S, 23° W) to 0.6 nM (56° S, 15° W). Enhanced concentrations of iron over the shelf may explain enhanced productivity around the South Sandwich Islands²¹ and the Galapagos Islands⁸.

The strong maxima in iron and phytoplankton, and the consequent CO₂ undersaturation, at the PF relative to the ACC (Fig. 4) show that the iron hypothesis holds for this unperturbed ecosystem. In contrast, although similar maxima for iron and chl *a* were observed in a natural plume downstream of the Galapagos Islands^{8,9}, significant effects of CO₂ were not observed. Moreover, in the equatorial Pacific experiments^{8,9}, pulsed addition of iron caused only a small and transient CO₂ response, whereas the natural continuous supply of iron in the PF produces a strong response.

The ACC southern part of the section (50–56° S) or 'ACC-proper' is truly offshore, well away from any shelf source influence. Here we compare the ACC with the subarctic North Pacific Ocean (Table 1). For the ACC, three factors favour stronger (more than five times) input from below: (1) the twofold higher wind stress²² which drives Ekman pumping upwelling; (2) the fourfold less density stratification (Fig. 3a), hence higher vertical mixing coefficient K_z ; (3) the twofold higher (Fig. 3b) deep-water Fe concentration, as compared to the subarctic North Pacific Ocean. Yet input from above by aerosols appears to be

FIG. 3 Comparison of the Antarctic circumpolar current (solid line or filled symbols) and the subarctic Pacific (dotted line, open symbols) for: *a*, potential density as a function of depth at 56° S, 12° W in the ACC, and at 50° N, 145° W at ocean station Papa³⁵. *b*, Dissolved Fe in the complete water column at 56° S, 12° W in the ACC and at 50° N, 145° W at ocean station Papa³⁵.



ten times less in the ACC than in the subarctic North Pacific. The relative dominance of above and below inputs is exactly the opposite in these two regions (Table 1). Recently for another high nutrient/low chlorophyll (HNLC) region, the equatorial Pacific, dominant input from below^{23,24} rather than above²⁵ was also demonstrated. The minor annual ACC aerosol input may accumulate in the winter sea-ice cover (Table 1). On melting of sea ice in the spring, the sudden input of iron to the upper water column (Table 1) might contribute to development of marginal ice zone (MIZ) blooms, unless within 2–4 days after melting storm events mix both iron and algae deep down and blooming ends prematurely²⁶. Storms (winds $>30 \text{ m s}^{-1}$) were encountered weekly during the eight weeks of the expedition and MIZ blooms

did not occur. Even with perfect stratification and insolation, pristine MIZ blooms may succumb due to catastrophic²⁶ grazing events.

The iron requirements of oceanic algae are not yet well known; published values range from $\text{Fe}:\text{C} > 2 \times 10^{-5}$ for prokaryotic cyanobacteria (*Synechococcus* sp. A2169)²⁷ to as low as $(\text{Fe}:\text{C}) \sim 2 \times 10^{-6}$ for a eukaryotic alga²⁸. When assuming a ratio of $\text{Fe}:\text{C}$ of 10^{-5} for oceanic algae, the ACC can sustain production of $167 \text{ mg C m}^{-2} \text{ d}^{-1}$, which compares well with the range of $80\text{--}300 \text{ mg C m}^{-2} \text{ d}^{-1}$ measured¹⁸ along the ACC part of section 11 (low compared to aforementioned $1,200\text{--}3,000 \text{ mg C m}^{-2} \text{ d}^{-1}$ in the PF). When assuming the measured recycling efficiency (*f*-ratio) (based on ¹⁵N assessment; F.

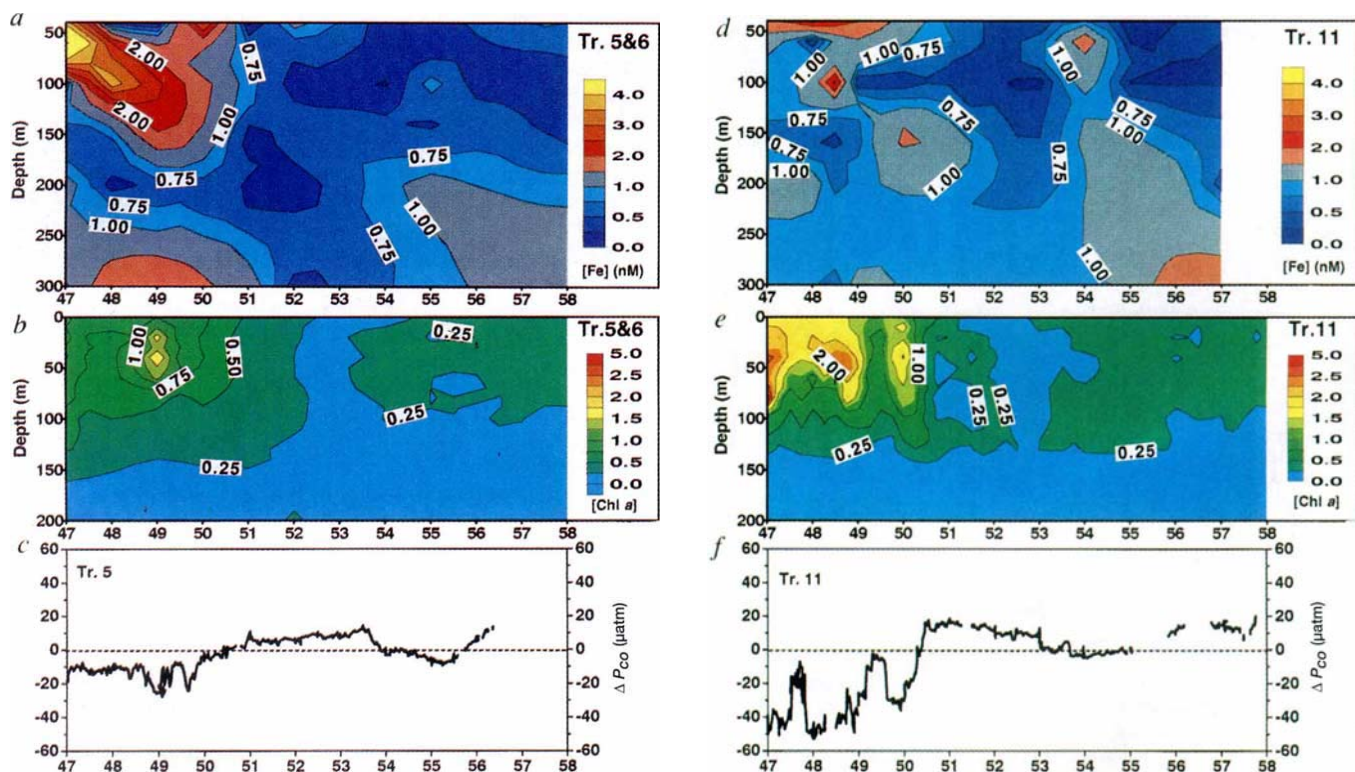


FIG. 4 *a*, Section plot of dissolved Fe (nM) at 40–300 m depth along the 6° W meridian from 47° to 58° S observed in the 24 October–6 November 1992 period. Mean value for the upper 120 m is 0.49 nM in ACC and 1.87 nM in PF. *b*, Section plot of chlorophyll *a* ($\mu\text{g per kg}$ sea water) at 0–200 m depth along the 6° W meridian from 47° to 58° S observed in the 24 October–6 November 1992 period. Mean value for upper 120 m is $0.25 \mu\text{g kg}^{-1}$ in ACC and $0.74 \mu\text{g kg}^{-1}$ in PF (inventories are 30 and $89 \mu\text{g m}^{-2}$ respectively). *c*, Plot of difference in

CO_2 partial pressure (P_{CO_2}) between surface waters (10 m depth) and marine air at the same section observed in the 24 October–6 November 1992 period. *d*, As *a* but for the period 10–21 November 1992. Mean value of dissolved iron for the upper 120 m is 0.31 nM in the ACC, and 1.14 nM in PF. *e*, As *b* but for the period 10–21 November 1992. Mean value of [chl *a*] for the upper 120 m is $0.24 \mu\text{g kg}^{-1}$ in the ACC, and $1.33 \mu\text{g kg}^{-1}$ in PF (inventories are 29 and $160 \mu\text{g m}^{-2}$ respectively). *f*, As *c*, but for the period 10–21 November 1992.

Dehairs, personal communication) of ~ 0.5 to apply for iron as well, the sustained primary production might be twice as high at $\sim 334 \text{ mg C m}^{-2} \text{ d}^{-1}$. In fact the f -ratio for different nutrient elements N and Fe may not be the same, but this requires new experimental approaches for investigation^{24,29}.

The upwelling supply of nitrate is $4.54 \times 10^{-6} \text{ mol m}^{-2} \text{ d}^{-1}$, which from the Redfield stoichiometry C:N=106:16 and an f -ratio of ~ 0.5 would support primary production of $720 \text{ mg C m}^{-2} \text{ d}^{-1}$. Comparison with the above actual production of $80\text{--}300 \text{ mg C m}^{-2} \text{ d}^{-1}$ shows that only part of the nitrate supply is thus utilized as its demand is restricted by the other growth limitations (iron, light, grazing).

In the relatively warm polar-front waters melting icebergs may further contribute Fe, thus further enhancing phytoplankton specific growth rates¹⁵. Seawater samples were collected in the wake of a 110-m high iceberg at 48°S , 06°W . This wake was of low salinity, rich in ice debris, and fizzing because of gas that had been trapped in the ice. These samples contained 2–9 nM Fe (unfiltered) as compared to $\sim 1 \text{ nM}$ at nearby ice-free stations and 20.4 nM in a large chunk of iceberg debris. Local nanoplankton abundance and total chlorophyll *a* around the iceberg were occasionally, but not significantly, higher than at the iceberg-free station. Also, when ascribing the slightly lower salinity of the whole PF region (~ 33.6 compared to ~ 33.8 in the $52\text{--}56^\circ \text{S ACC}$) to iceberg melting, the ensuing iron input would be insignificant because of massive dilution.

Blooms commonly have a high f -ratio for nitrogen (and an assumed high f -ratio for iron) and a high export production of C, N, P and Fe. This would lead to the surface-water inventory of Fe rapidly becoming depleted in summer, as also suggested by the decreasing Fe inventory in the PF during spring (Fig. 4). Simultaneously, the upwelling and eddy-diffusion supply of Fe from below would have diminished as the wind stress decreases in summer²². In other words, wind stress plays an opposite role for the light-limited regime relative to the iron-limited regime. This may be alleviated by the notion that algal cells in optimum light regime (stratified summer conditions) are expected to have lower specific iron requirements³⁰. Without summer studies in offshore ACC waters, however, the occasional collapse of summer productivity solely due to iron stress cannot be ruled out.

On Fe enhancement, either naturally in neritic^{12–14} waters and the PF, or experimentally^{2,4–7,15,16} in incubation bottles, plankton doubling-rates increase and, provided that the light regime is adequate and grazing is modest, blooms will develop to some extent. Blooms are an important vehicle for export of carbon into the deep sea (the biological pump) and associated draw-down of CO_2 from the atmosphere. It is to be noted though that macronutrients are by no means fully utilized yet, whereas the biomass inventory of the blooms is still not optimal. In contrast with the iron-replete polar front, the low Fe levels in the 'ACC-proper' can only sustain its deep-mixed, poorly developed, retention³¹ ecosystem with little or no export, which is not able to cause undersaturation of CO_2 in surface waters. This very deeply mixed ($>100 \text{ m}$) system still constitutes a significant biomass inventory ($30 \mu\text{g chl } a \text{ m}^{-2}$; see Fig. 4 legend) but largely escapes satellite observation. □

Received 27 May; accepted 13 December 1994.

- Gran, H. H. *Rapp. P.-v. Cons. perm. int. Explor. Mer* **75**, 37–46 (1931).
- Martin, J. H. & Fitzwater, S. E. *Nature* **331**, 341–343 (1988).
- de Baar, H. J. W. *Prog. Oceanogr.* **33**, 347–386 (1994).
- Coale, K. H. *Limnol. Oceanogr.* **36**, 1851–1864 (1991).
- de Baar, H. J. W. *et al. Mar. Ecol. Prog. Ser.* **65**, 105–122 (1990).
- Buma, A. G. J., de Baar, H. J. W., Nolting, R. F. & van Bennekom, A. J. *Limnol. Oceanogr.* **36**, 1865–1878 (1991).
- Martin, J. H. & Gordon, R. M. *Globi Biogeochem. Cycles* **4**, 5–12 (1990).
- Martin, J. H. *et al. Nature* **371**, 123–129 (1994).
- Watson, A. J. *et al. Nature* **371**, 143–145 (1994).
- Dugdale, R. C. & Wilkerson, F. P. *Globi Biogeochem. Cycles* **4**, 13–19 (1990).
- Banse, K. *Limnol. Oceanogr.* **36**, 1886–1898 (1991).
- Hart, T. J. *Discovery Rep.* **8**, 1–268 (1934); *Discovery Rep.* **21**, 261–365 (1942).
- Westerlund, S. & Ohman, P. *Mar. Chem.* **35**, 199–217 (1991).
- Nolting, R. F., de Baar, H. J. W., van Bennekom, A. J. & Masson, A. *Mar. Chem.* **35**, 219–243 (1991).

- van Leeuwe, M. A., Scharek, R., de Baar, H. J. W., de Jong, J. T. F. & Goeyens, L. *EOS (abstr.)* **75**, 178 (1994).
- Scharek, R., van Leeuwe, M. A., de Baar, H. J. W., de Jong, J. T. F. & Goeyens, L. *EOS (abstr.)* **75**, 178 (1994).
- Bathmann, U., Smetacek, V., de Baar, H. J. W., Fahrbach, E. & Krause, G. (eds.) *Ber. Polarforsch. Schifff.* **135**, 4–126 (1994).
- Jochem, F. J., Mathot, S. & Quéguiner, B. *EOS (abstr.)* **75**, 158 (1994).
- de Baar, H. J. W. *et al. EOS (abstr.)* **75**, 178 (1994).
- Martin, J. H., Gordon, R. M. & Fitzwater, S. E. *Nature* **345**, 156–158 (1990).
- Perisintono, R., Laubscher, R. K. & McQuaid, C. D. *Mar. Ecol. Prog. Ser.* **88**, 41–53 (1992).
- Trenberth, K. E., Large, W. G. & Olson, J. G. *J. Phys. Oceanogr.* **20**, 1742–1760 (1990).
- Gordon, M., Coale, K. & Johnson, K. *EOS (abstr.)* **75**, 114 (1994).
- Wells, M. L., Price, N. M. & Bruland, K. W. *Mar. Chem.* (in the press).
- Martin, J. H. *Nature* **353**, 123 (1991).
- Lancelot, C., Mathot, S., Veth, C. & de Baar, H. J. W. *Polar Biol.* **13**, 377–387 (1993).
- Brand, L. *Limnol. Oceanogr.* **36**, 1756–1771 (1991).
- Sunda, W. G., Swift, D. & Huntsman, S. A. *Nature* **351**, 55–57 (1991).
- Hutchins, D. A., DiTullio, R. G., Bruland, K. W. *Limnol. Oceanogr.* **38**, 1242–1255 (1993).
- Raven, J. A. *New Phytol.* **116**, 1–18 (1990).
- Smetacek, V., Scharek, R. & Nöthig, E. M. in *Antarctic Ecosystems* (eds Kerry, K. R. & Hempel, G.) 103–114 (Springer, Berlin, 1990).
- Gordon, A. L., Taylor, H. W. & Georgi, D. T. in *Polar Oceans* (ed. Dunbar, M. J.) 45–67 (Arctic Inst. N. Am., McGill Univ., Montreal, 1977).
- Miller, C. B., Frost, B. W., Wheeler, P. A., Landry, M. R., Welschmeyer, N. & Powell, T. M. *Limnol. Oceanogr.* **36**, 1600–1615 (1991).
- Gargett, A. E. *Limnol. Oceanogr.* **36**, 1527–1545 (1991).
- Martin, J. H., Gordon, R. M., Fitzwater, S. E. & Broenkow, W. W. *Deep-Sea Res.* **35**, 649–680 (1989).
- Broecker, W. S. in *Evolution of Physical Oceanography* (eds Warren, B. A. & Wunsch, C. A.) 460 (MIT Press, Cambridge, Massachusetts, 1981).
- Olbers, D. & Wenzel, M. in *Oceanic Circulation Models: Combining Data and Dynamics* (eds Anderson, D. L. T. & Willebrand, J.) 95–139 (Kluwer Academic, Dordrecht, 1989).
- Garcon, V. C., Thomas, F., Wong, C. S. & Minster, J. F. *Deep-Sea Res.* **39**, 921–938 (1986).
- Duce, R. A. & Tindale, N. W. *Limnol. Oceanogr.* **36**, 1715–1726 (1991).
- Nolting, R. F., de Jong, J. T. M., Löscher, B. M. & de Baar, H. J. W. *EOS (abstr.)* **75**, 219 (1994).
- Timmermans, K. R., Gledhill, M., van den Berg, C. S., Nolting, R. F. & de Baar, H. J. W. *EOS (abstr.)* **75**, 78 (1994).

ACKNOWLEDGEMENTS. We thank the officers, engineers and crew of RV *Polarstern* for their support. This work was supported by the Commissie Antarctisch Onderzoek, the Netherlands Climate Research programme, the Netherlands Organisation for Scientific Research (NWO), the German Ministry for Research and Technology (BMFT) and the Alfred Wegener Institute for Polar and Marine Research, which made the *Polarstern* available to us.

A silicate weathering mechanism linking increases in marine $^{87}\text{Sr}/^{86}\text{Sr}$ with global glaciation

Joel D. Blum* & Yigal Erel†*

* Department of Earth Sciences, Dartmouth College, Hanover, New Hampshire 03755, USA

† Institute of Earth Sciences, Hebrew University, Jerusalem 91909, Israel

THE $^{87}\text{Sr}/^{86}\text{Sr}$ ratio of sea water is generally regarded as a proxy for the rate of chemical weathering of the continents. Interpretation of the $^{87}\text{Sr}/^{86}\text{Sr}$ record in marine limestones requires an understanding of how changes in this quantity are linked to climate and tectonic events. A connection between the strontium isotope record and the extent of global glaciation has been proposed^{1–8}, but without a well established mechanism. Glaciation may influence marine $^{87}\text{Sr}/^{86}\text{Sr}$ by changing global riverine Sr fluxes (that is, rock weathering rates) and/or $^{87}\text{Sr}/^{86}\text{Sr}$ ratios (that is, relative mineral weathering rates). Here we investigate how the $^{87}\text{Sr}/^{86}\text{Sr}$ release from silicate weathering is affected by glaciation, using the Sr isotope systematics of granitoid soils on alpine glacial moraines in the Wind River Range, Wyoming. We observe a negative correlation between the $^{87}\text{Sr}/^{86}\text{Sr}$ ratio of exchangeable Sr in soils and the soil age, indicating that the $^{87}\text{Sr}/^{86}\text{Sr}$ ratio of Sr released in the early stages of weathering is significantly higher than in later stages. We estimate that this mechanism can increase global riverine $^{87}\text{Sr}/^{86}\text{Sr}$ by an average of 0.0002 during periods of glacial-interglacial cycling. When superimposed on long-term trends controlled by tectonic processes^{2–7,9,10}, this shift can account for the correlation between the rate of change of marine $^{87}\text{Sr}/^{86}\text{Sr}$ and the intensity of glaciation over at least the past 10 Myr.

TABLE 1 Rb and Sr data from soil, water and snow

Sample	Map unit ^{13,14}	Age (kyr)	Ammonium-acetate extractable			Whole-soil digestion		
			⁸⁷ Sr/ ⁸⁶ Sr (±2σ)	Rb (μg g ⁻¹)	Sr (μg g ⁻¹)	⁸⁷ Sr/ ⁸⁶ Sr (±2σ)	Rb (μg g ⁻¹)	Sr (μg g ⁻¹)
AT-FL-5H	Sacagawea Ridge till	≥297	0.711369±23	0.29	19.86	0.712966±40	65	407
AT-FL-18E	Sacagawea Ridge gravel	≥297	0.711619±24	0.24	7.19	0.717969±43	79	354
AT-TL-7G	Bull Lake till	96–133	0.716058±28	0.73	28.52	0.710341±25	59	503
AT-TL-14G	Pinedale till	19.5–22.5	0.718665±39	0.15	7.43	0.714580±20	53	393
AT-TL-28D	Titcomb Lake till	10.4–11.4	0.723822±24	0.07	1.22	0.726378±26	50	292
AT-TL-29C	Audubon till	1.5–2.5	0.735523±29	0.49	1.14	0.727230±36	57	284
AT-TL-23B	Gannett Peak till	0.3–0.5	0.794667±27	0.40	2.65	0.725129±23	66	346
AT-TL-31	Stream water	—	0.738530±26	0.30*	2.41*			
AT-TL-34	Stream water	—	0.743111±29	0.35*	2.37*			
	Wind River Range snowpack (Union Pass)	(1992)†	0.710643±35	0.017*	0.13*			

The Sacagawea Ridge, Bull Lake and Pinedale till soil samples are from terminal moraines at 2,300–2,600 m above sea level (a.s.l.) mapped by Richmond¹³. The Titcomb Lake^{15,16}, Audubon and Gannett Peak till soil samples are from terminal moraines at 3,200–3,350 m a.s.l. mapped by Mahaney *et al.*¹⁴. The water samples are from a stream that drains a valley last glaciated by the Audubon glacial advance, which also contains a small area of Gannett till and small active glaciers. The very low concentration of Sr in the snowpack sample shows the small contribution of atmospheric Sr to the high-elevation streamwater samples. Ages of glacial advances are from ¹⁰Be exposure ages^{15,16}, except for the Audubon which is estimated from correlations with other glacial deposits dated¹⁴ by ¹⁴C, and the Gannett Peak which is based on the historic record of 'Little Ice Age' global glacial advances. Soils were collected from B-horizons at 30–40 cm depths in soil pits dug on the crests of terminal moraines, all within the Fremont Creek drainage basin in the northern Wind River Range 2–35 km north of Pinedale, Wyoming, USA. Soil samples (~2 kg) were dried at 50 °C and sieved to pass a 2-mm mesh. Representative sub-samples were (1) exchanged with ultrapure ammonium acetate buffered at pH 7 and the extract separated from the sample by centrifugation, and (2) digested in ultrapure HF and HClO₄. Extracts and digests were analysed for Sr isotopic composition, and Sr and Rb concentration, by isotope dilution thermal ionization mass spectrometry following procedures given in ref. 8. Uncertainties in ⁸⁷Sr/⁸⁶Sr are 2σ in the last two decimal places. Uncertainties in Rb and Sr concentrations are ~1%. Whole-soil digestion ⁸⁷Sr/⁸⁶Sr ratios decrease in the oldest soils as expected as biotite is weathered from the parent material.

* These concentrations in ng g⁻¹, not μg g⁻¹.

† This sample represents a composite sample of the snowpack from the winter of 1992–93.

By studying stream drainages with varying glacial histories we demonstrated previously⁸ that accelerated weathering of biotite following deglaciation led to elevated riverine ⁸⁷Sr/⁸⁶Sr. To evaluate the global implications of this mechanism we investigated the isotopic composition of Sr released by weathering of six glacial moraines with ages varying from ~0.4 to ≥297 kyr (Table 1) in the Fremont Creek drainage basin in the northern Wind River Range. This area was chosen because bedrock consists of Precambrian granitoids and gneisses typical of continental shield rocks, for which the Rb–Sr mineral system was last

disturbed ~2.0 Gyr ago¹¹ (close to the ~1.5 Gyr average continental crust Rb–Sr mineral age)¹². In addition, the area has Quaternary glacial deposits and soils that have previously been mapped and dated^{13–16}. By studying Sr release in soils of variable age developed on a single parent material, we have established a record of how the isotopic composition of Sr released by silicate weathering varies with time following the exposure of fresh rock by deglaciation.

A dramatic and systematic decrease is observed in the ⁸⁷Sr/⁸⁶Sr ratio of exchangeable Sr with increasing soil age ranging from 0.79467 down to 0.71137 (Fig. 1). These shifts are far in excess of the observed variability in ⁸⁷Sr/⁸⁶Sr of total digestions of the same samples (Table 1), and are believed to be the result of accelerated biotite weathering in the younger soils. Biotite is a common mineral in the parent material with ⁸⁷Sr/⁸⁶Sr ≥0.92450 (ref. 11), and is capable of elevating the ⁸⁷Sr/⁸⁶Sr ratios dramatically in young soils⁸. Decreases in biotite and increases in vermiculite (a common biotite weathering product) have been observed with increasing age in the post-Wisconsin soils studied here¹⁴, and on the Canadian Shield¹⁷. The ⁸⁷Sr/⁸⁶Sr of stream waters draining a valley glaciated by the Audubon glacial advance (~2 kyr ago) are consistent with the values for soil-exchangeable Sr in the Audubon moraines (Fig. 1). The shift in ⁸⁷Sr/⁸⁶Sr between the 11 and ~115 kyr soils in the Wind River Range is consistent with a previous study in the Sierra Nevada⁸ when scaled to account for the 20-fold difference in bedrock ages.

To compare the shifts in ⁸⁷Sr/⁸⁶Sr of the soil from the Wind River Range with changes in global riverine ⁸⁷Sr/⁸⁶Sr, we linearly extrapolated between our data points and calculated the average weathering-derived ⁸⁷Sr/⁸⁶Sr in the Wind River Range for the time periods 0–0.4, 0.4–2, 2–11, 11–21 and 21–100 kyr following glaciation. Next we adopted the 100-kyr periodicity for major glacial advances (observed for the Quaternary) and calculated

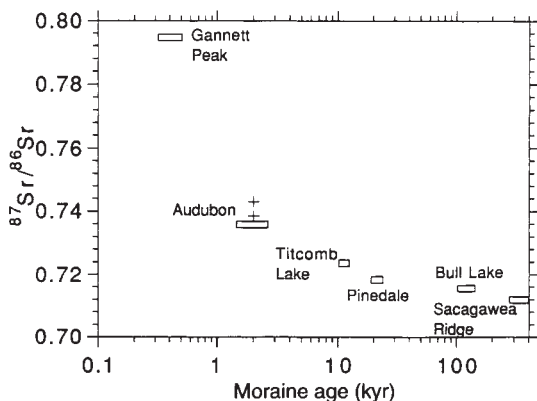


FIG. 1 The ⁸⁷Sr/⁸⁶Sr ratio in the ammonium-acetate extractable fraction of B-horizon soils developed on glacial moraines plotted against the log of the age of the moraines. Also plotted (crosses) is the ⁸⁷Sr/⁸⁶Sr ratio of two stream water samples draining glacial deposits of predominantly Audubon till. Rectangular symbols indicate uncertainties in moraine ages and are larger than uncertainties in ⁸⁷Sr/⁸⁶Sr.

the difference in riverine $^{87}\text{Sr}/^{86}\text{Sr}$ between that for each of the younger time intervals above and that at 100 kyr ago. Although the youngest soil has extremely high $^{87}\text{Sr}/^{86}\text{Sr}$, it is so young that it contributes only a small fraction of the time-integrated riverine $^{87}\text{Sr}/^{86}\text{Sr}$ elevation over the glacial-interglacial timescale considered here (Fig. 2).

If these results can be considered to be globally representative (and we believe that this should be so, although it remains to be demonstrated), we can assess the implications for global riverine $^{87}\text{Sr}/^{86}\text{Sr}$ by considering what portion of the global Sr flux comes from areas affected by glaciation. We have taken two approaches to this estimation, which give essentially the same result. First, we assumed that the Sr flux scales with land area, and adopted a value of 20% of the global Sr flux because this is the portion of the world's land area that was glaciated during the Last Glacial Maximum¹⁸ (this disregards the additional ~10% land area that was covered by fluvial and aeolian glacial deposits)¹⁹. Second, we used the dissolved Sr flux and drainage areas from the 39 major Canadian rivers²⁰ to calculate an average Sr flux of $220 \text{ mol Sr yr}^{-1} \text{ per km}^2$ of surface area for Canada. This value was multiplied by the global area affected by glaciation ($28.3 \times 10^6 \text{ km}^2$)¹⁸ to arrive at a global Sr flux from glaciated areas of $6.23 \times 10^9 \text{ mol yr}^{-1}$, which is 19% of the $3.33 \times 10^{10} \text{ mol yr}^{-1}$ global riverine dissolved Sr flux¹⁰. An additional parameter that must be estimated is the proportion of the Sr flux derived from silicate rather than carbonate weathering, which has been calculated to average 40% globally²¹, but for which we have adopted a more conservative estimate of 20% calculated only for the Canadian Shield²⁰.

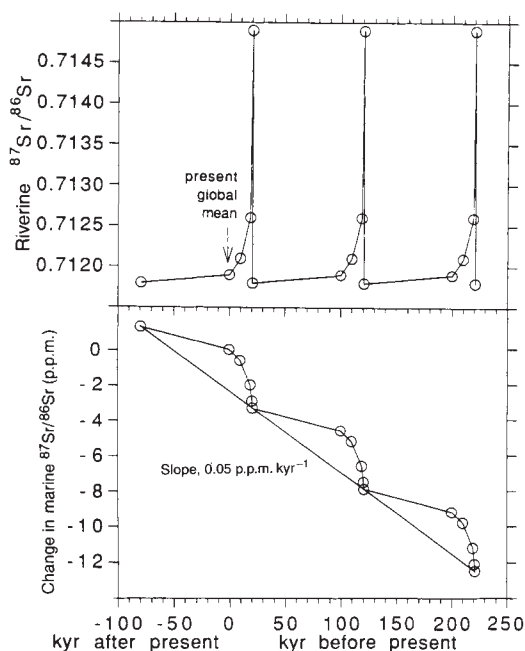


FIG. 2 Plot of the calculated global average riverine $^{87}\text{Sr}/^{86}\text{Sr}$ (top) and the change in the marine $^{87}\text{Sr}/^{86}\text{Sr}$ (bottom) against time (in kyr before and after the present), showing changes caused by 100-kyr glacial-interglacial cycles, based on the assumption that each cycle affects riverine Sr similarly and can be estimated based on our data from the Wind River Range (Fig. 1). The calculated change in marine $^{87}\text{Sr}/^{86}\text{Sr}$ assumes that before glaciation the ocean was in a steady state with respect to $^{87}\text{Sr}/^{86}\text{Sr}$, that all Sr fluxes remain constant, and that riverine $^{87}\text{Sr}/^{86}\text{Sr}$ varies as shown. Note that the sharp spike in the riverine $^{87}\text{Sr}/^{86}\text{Sr}$ caused by the most radiogenic sample (the Gannett Peak till soil) represents a very short time interval and therefore this single sample has very little effect on the marine $^{87}\text{Sr}/^{86}\text{Sr}$ averaged over 100-kyr glacial cycles. The average shift in global riverine $^{87}\text{Sr}/^{86}\text{Sr}$ is 0.0002 and the rate of change of marine $^{87}\text{Sr}/^{86}\text{Sr}$ is $0.05 \text{ p.p.m. kyr}^{-1}$ when averaged over a 100-kyr cycle.

Our best estimate of how to 'correct' the $^{87}\text{Sr}/^{86}\text{Sr}$ shifts observed in the Wind River Range data to assess the influence of weathering on global riverine $^{87}\text{Sr}/^{86}\text{Sr}$ is then to multiply the shifts by 0.04 ($20\% \times 20\%$) to correct for the factors described above. Finally we calculated how the global riverine $^{87}\text{Sr}/^{86}\text{Sr}$ ratio might have varied with time on a glacial-interglacial timescale by fixing the global riverine $^{87}\text{Sr}/^{86}\text{Sr}$ at the present-day value (0.7119) 21 kyr after the glacial maximum, and then superimposing the shifts in $^{87}\text{Sr}/^{86}\text{Sr}$ estimated from the Wind River Range (Fig. 2). Just before deglaciation the average riverine value is estimated to have been 0.7118, and immediately following deglaciation the value would have risen to 0.7149 and then rapidly fallen over 21 kyr to 0.7119 (Fig. 2), giving an average increase in global riverine $^{87}\text{Sr}/^{86}\text{Sr}$ of 0.0002 during periods of glacial-interglacial cycling. Obviously, neither glaciation nor deglaciation is an instantaneous event, but they are modelled as such here to achieve a first-order estimate of the effect this mechanism could have on marine $^{87}\text{Sr}/^{86}\text{Sr}$.

To evaluate the potential influence that glaciation may have had on the marine $^{87}\text{Sr}/^{86}\text{Sr}$ ratio, we use the general equation relating the Sr flux and $^{87}\text{Sr}/^{86}\text{Sr}$ ratio of riverine (continental crust), hydrothermal (oceanic crust) and deep-sea carbonate (diagenesis and dissolution) input to the ocean to the rate of change of the seawater $^{87}\text{Sr}/^{86}\text{Sr}$ ratio^{4,6,9,10}, and adopt values from the literature for the various Sr reservoirs¹⁰. We hold constant the riverine Sr flux in our calculations to isolate the effect of changing riverine $^{87}\text{Sr}/^{86}\text{Sr}$, although it has been suggested that glaciation may also increase Sr fluxes by exposing fresh mineral surfaces¹⁻⁷, which would slightly enhance the shifts in marine $^{87}\text{Sr}/^{86}\text{Sr}$ that we calculate. The results of these calculations are shown in Fig. 2, and demonstrate that the silicate weathering mechanism can produce shifts in marine $^{87}\text{Sr}/^{86}\text{Sr}$ of up to 0.000003 (3 p.p.m.) in the first 21 kyr after deglaciation and shifts of up to $0.05 \text{ p.p.m. kyr}^{-1}$ when averaged over a 100-kyr glacial-interglacial cycle.

Next we compare our estimates with the marine $^{87}\text{Sr}/^{86}\text{Sr}$ record and global glacial history derived from the marine $\delta^{18}\text{O}$ record over the past 10 Myr, the time period for which high-resolution marine isotopic records are available. Following a period of almost no change in marine $^{87}\text{Sr}/^{86}\text{Sr}$ between 10 and 6 Myr ago, marine $^{87}\text{Sr}/^{86}\text{Sr}$ increased by $0.08 \text{ p.p.m. kyr}^{-1}$ between 6.2 and 4.5 Myr ago⁵, which was a period corresponding to glacial development in the Southern Hemisphere²². This was followed by a Southern Hemisphere warm period between 3.4 and 2.4 Myr (ref. 23) and correspondingly no change in marine $^{87}\text{Sr}/^{86}\text{Sr}$ between 4.5 and 2.5 Myr ago⁵. The onset of low-amplitude Northern Hemisphere glaciation 2.4 Myr ago²⁴ coincided almost exactly with an increase in marine $^{87}\text{Sr}/^{86}\text{Sr}$ at a rate of $\sim 0.05 \text{ p.p.m. kyr}^{-1}$ (ref. 6), and the transition to higher-amplitude Northern Hemisphere glaciation between 0.8 Myr and the present²⁵ corresponded to an even steeper rise in marine $^{87}\text{Sr}/^{86}\text{Sr}$ at a rate of $0.10 \text{ p.p.m. kyr}^{-1}$ (ref. 6). Thus major changes in the Earth's glacial history over the past 10 Myr do correlate qualitatively with changes²⁻⁷ in marine $^{87}\text{Sr}/^{86}\text{Sr}$. The magnitude of $^{87}\text{Sr}/^{86}\text{Sr}$ changes correlated with glaciations (0.05 – $0.10 \text{ p.p.m. kyr}^{-1}$) are, however, in some cases greater than the $\sim 0.05 \text{ p.p.m. kyr}^{-1}$ shift predicted by our proposed mineral weathering mechanism. Changes in marine $^{87}\text{Sr}/^{86}\text{Sr}$ on the time-scale of individual glacial-interglacial cycles are smaller than measurement uncertainties^{26,27}.

Although the resolution of the $>10 \text{ Myr}$ marine isotope record does not allow detailed comparison between glacial intensity and the fine structure of marine $^{87}\text{Sr}/^{86}\text{Sr}$, it has also been pointed out that six pulses of glaciation between 10 and 35 Myr ago seem to be correlated with increases in the slope of the $^{87}\text{Sr}/^{86}\text{Sr}$ record^{7,28}. Nevertheless, these correlations do not prove causality, and compelling arguments have been made that some features of the marine $^{87}\text{Sr}/^{86}\text{Sr}$ record may also correlate with changes in uplift rates and/or monsoonal circulation associated with uplift of the Himalayan plateau^{3,10,28,29} or that tectonic

uplift may have triggered global glaciations³⁰. As suggested by others, it is our view that glaciation is responsible for the fine structure superimposed on long-term tectonically driven trends in the $^{87}\text{Sr}/^{86}\text{Sr}$ record^{28,29} and that increases in riverine $^{87}\text{Sr}/^{86}\text{Sr}$ related to enhanced biotite weathering following glaciations is a plausible mechanism linking glaciations with the marine $^{87}\text{Sr}/^{86}\text{Sr}$ record. □

Received 27 May; accepted 15 December 1994.

1. Armstrong, R. L. *Nature* **230**, 132–133 (1971).
2. DePaolo, D. J. *Geology* **14**, 103–106 (1986).
3. Raymo, M. E., Ruddiman, W. F. & Froelich, P. N. *Geology* **16**, 649–653 (1988).
4. Hodell, D. A., Mueller, P. A., McKenzie, J. A. & Mead, G. A. *Earth planet. Sci. Lett.* **92**, 165–178 (1989).
5. Hodell, D. A., Mead, G. A. & Mueller, P. A. *Chem. Geol.* **80**, 291–307 (1990).
6. Capo, R. C. & DePaolo, D. J. *Science* **249**, 51–55 (1990).
7. Miller, K. G., Feigenson, M. D., Wright, J. D. & Clement, B. M. *Paleoceanography* **6**, 33–52 (1991).
8. Blum, J. D., Erel, Y. & Brown, K. *Geochim. cosmochim. Acta* **58**, 5019–5025 (1993).
9. Richter, F. M. & Turekian, K. K. *Earth planet. Sci. Lett.* **119**, 121–131 (1993).
10. Palmer, M. R. & Edmond, J. M. *Earth planet. Sci. Lett.* **92**, 11–26 (1989).
11. Naylor, R. S., Steiger, R. H. & Wasserburg, G. J. *Geochim. cosmochim. Acta* **34**, 1133–1159 (1970).
12. Hurley, P. M. & Rand, J. R. *Science* **164**, 1229–1242 (1969).

13. Richmond, G. M. *Quad. Map GQ-1138*, scale 1:24,000 (US Geol. Surv., Washington DC, 1973).
14. Mahaney, W. C. & Halvorson, D. L. in *Rates of Chemical Weathering of Rocks and Minerals* (eds Colman, S. M. & Dethier, D. P.) 147–167 (Academic, Orlando, 1986).
15. Gosse, J., Klein, J., Evenson, E., Lawn, B. & Middleton, R. *Circ. No. 1107* 114 (US Geol. Surv., Denver, Colorado, 1994).
16. Evenson, E., Klein, J., Lawn, B., Middleton, R. & Gosse, J. *Geol. Soc. Am. Abstr. Progr.* **26**, A-511 (1994).
17. Ross, G. J., Wang, C. & Protz, R. in *Correlation of Quaternary Chronologies* (ed. Mahaney, W. C.) 191–202 (Geobooks, Norwich, UK, 1983).
18. Hughes, T. J. et al. in *The Last Great Ice Sheets* (eds Denton, G. H. & Hughes, T. J.) 275–318 (Wiley, New York, 1981).
19. Taylor, S. R., McLennan, S. M. & McCulloch, M. T. *Geochim. cosmochim. Acta* **47**, 1897–1905 (1983).
20. Wadleigh, M. A., Veizer, J. & Brooks, C. *Geochim. cosmochim. Acta* **49**, 1727–1736 (1985).
21. Bickle, M. J. *Nature* **367**, 699–704 (1994).
22. Kennett, J. P. *Init. Rep. DSDP Leg. 90*, 1388–1412 (1986).
23. Malmgren, B. A. & Berggren, W. A. *Paleoceanography* **2**, 445–456 (1987).
24. Shackleton, N. J. et al. *Nature* **307**, 620–623 (1984).
25. Kennett, J. P. & von der Borch, C. *Init. Rep. DSDP Leg. 90*, 1493–1517 (1986).
26. Henderson, G. M., Martel, D. J., O'Nions, R. K. & Shackleton, N. J. *Mineral. Mag.* **58A**, 404–405 (1994).
27. Clemens, S. C., Gromet, P. & Farrell, J. W. *Mineralog. Mag.* **58A**, 178–179 (1994).
28. Oslick, J. S., Miller, K. G., Feigenson, M. D. & Wright, J. D. *Paleoceanography* **9**, 427–443 (1994).
29. Hodell, D. A. & Woodruff, F. *Paleoceanography* **9**, 405–426 (1994).
30. Ruddiman, W. F. & Kutzbach, J. E. *J. geophys. Res.* **94**, 18409–18427 (1989).

ACKNOWLEDGEMENTS. We thank M. Hingston and A. Taylor for their invaluable work in the laboratory and in the field. This work was supported by the US NSF and an Alfred P. Sloan Foundation fellowship.

The Oak Ridge fault system and the 1994 Northridge earthquake

Robert S. Yeats & Gary J. Huftle

Department of Geosciences, Oregon State University, Corvallis, Oregon 97331, USA

THE 17 January 1994 Northridge earthquake (moment magnitude $M_w = 6.7$) in the San Fernando Valley in southern California illuminated a hitherto unrecognized blind reverse fault with a moderate southward dip¹. This fault lies beneath the active north-dipping Santa Susana fault system and uplifted both the footwall and the hanging wall of the Santa Susana fault during the earthquake¹. Here we argue that footwall uplift on the Santa Susana fault before the earthquake could have been used to identify the Northridge blind thrust as an active fault. Moreover, we propose that the Northridge earthquake occurred on a continuation of the Oak Ridge fault system, which reaches the surface in the Ventura basin to the west. Slip rates on the western part of this fault system^{2,3} are nearly three times larger than on the Northridge blind thrust⁴, increasing the probability of a Northridge-sized earthquake in the heavily populated Ventura basin.

Since the 1971 San Fernando earthquake, many investigators have been mapping southern California faults to identify potential earthquake source zones that might produce destructive earthquakes in the Los Angeles metropolitan region. However, these detailed studies failed to identify in advance the south-dipping blind thrust that was the source of the 1994 Northridge earthquake¹. Was there evidence for the blind thrust that, in retrospect, could have led to its identification before the earthquake? Could such evidence be sought elsewhere in southern California and, indeed, in reverse-fault zones elsewhere in the world? Evidence presented here indicates that the blind thrust is part of the Oak Ridge fault system, which reaches the surface in the Ventura basin to the west and was previously identified in that area as a potential seismic source².

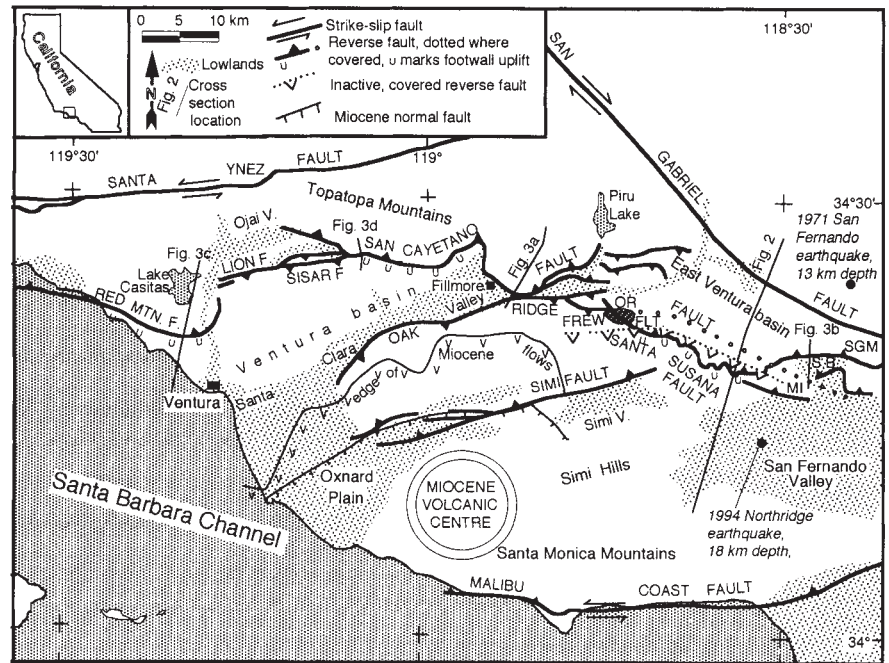
Although the Oak Ridge fault has low instrumental seismicity, and no major earthquakes have occurred in 200 years of record keeping, it has an estimated slip rate of about 5 mm yr⁻¹ (refs

2, 3). Satellite geodesy shows a shortening across the Ventura basin of 7–10 mm yr⁻¹, mostly taken up between the south-dipping Oak Ridge fault and the north-dipping Red Mountain and San Cayetano faults^{5,6}. This shortening must reflect elastic strain accumulation; it is being relieved neither by earthquakes nor by aseismic slip. Between these opposing fault systems, the Ventura basin itself (Fig. 1) contains a thick sequence of Pliocene and Pleistocene strata, including, south of Fillmore, the thickest Pleistocene section in the world (Fig. 1). This thick sequence is expressed as an east-trending 60-mGal gravity low⁷. The Oak Ridge fault extends eastward along the southern margin of the thick sequence and bifurcates, with one strand dying out to the east along the edge of the Santa Clara Valley and another strand extending into the Santa Susana Mountains beneath the north-dipping Santa Susana fault (Fig. 1)⁸. Until the Northridge earthquake, it had been assumed that the strand buried beneath the Santa Susana fault was inactive.

The gravity low continues uninterrupted beneath the Santa Susana Mountains into the San Fernando Valley⁷ and marks the position of the east Ventura basin and, farther east, the Sylmar basin in the northernmost San Fernando Valley. The southern edge of the gravity low has a convex-northward map pattern. This pattern is consistent with the curving trend of the Oak Ridge fault system, which we extend along the southern edge of the gravity low beneath the Santa Susana Mountains as a blind thrust. The adjacent, south-dipping Frew reverse fault is mapped using well data from the south side of Oak Ridge into the southern foothills of the Santa Susana Mountains, where it is overridden by the Santa Susana fault (Fig. 2), and into the San Fernando Valley (Fig. 1)⁸. The Frew fault, which is clearly inactive in that it is overlain unconformably by Pleistocene strata⁸, also follows a convex-north map pattern (Fig. 1) and is interpreted as an older member of the Oak Ridge fault system.

The Oak Ridge fault may have northward curvature because it is a reactivated Miocene normal fault, a ring fault related to emplacement of a large volcanic caldera in the western Santa Monica Mountains (Fig. 1)⁹. The caldera is bounded on the north by a Miocene normal fault in the Oxnard plain, partly reactivated by the Quaternary Simi reverse fault. Farther north, Miocene lava flows are buttressed against a series of Miocene uplifts parallel to and south of the Oak Ridge fault⁹. Both features show convex-northward curvature. Evidence that the Oak Ridge fault itself was active in the Miocene is found in the Oakridge oil field, where the fault separates a thin Miocene

FIG. 1 Tectonic map of north-dipping and south-dipping active reverse faults in the western Transverse Ranges, California. Solid triangles point in the direction of fault dip. Miocene features shown include caldera in western Santa Monica Mountains (concentric circles), large-displacement normal faults (solid line with hachures), and northern edge of Miocene lava flows (solid lines with v's). Oak Ridge fault dotted beneath the Santa Susana fault. Inactive Frew fault is dotted throughout its length because, throughout its length, it is overlain by Quaternary strata or by other faults. Mainshocks (dots) of 1971 and 1994 earthquakes are shown. OR, Oakridge oil field; MI, Mission Hills; SB, Sylmar basin; SGM, San Gabriel Mountains.



sequence on the south side from a thick Miocene sequence on the north side⁸.

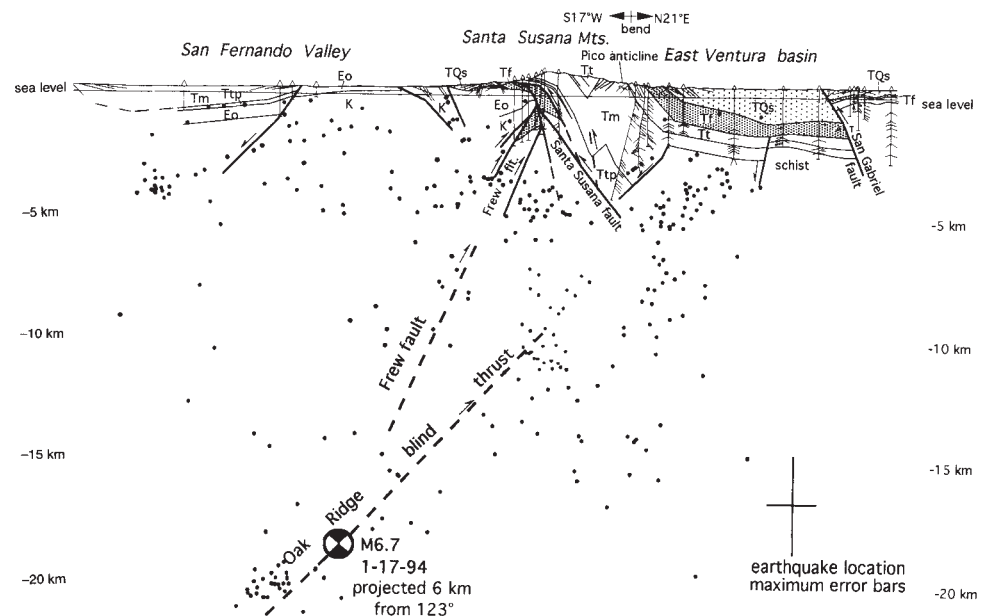
No primary surface rupture seems to have accompanied the Northridge earthquake, and the seismicity signature at shallow depths is diffuse¹, similar to blind-thrust earthquakes that struck Coalinga, California, in 1983 and Whittier Narrows, California, in 1987 (ref. 10). Diffuse patterns of seismicity and an absence of surface rupture also characterized the earthquakes in Niigata, Japan, in 1964 (ref. 11) and Puerto Limon, Costa Rica, in 1991 (ref. 12). The surface expression of each of these earthquakes is a broad anticlinal fold which grew during the earthquake. The diffuse pattern of seismicity in all these earthquakes is probably a response to internal deformation of the rocks as they are folded.

The Northridge earthquake differs from those cited above in that near-surface, north- and south-dipping faults are present, but these did not rupture the ground surface in 1994, although

they may have done so in the past. Like the examples cited above, the near-surface expression of the Northridge earthquake is also a broad fold, expressed by gentle southerly dips south of the Santa Susana Mountains and moderately steep northerly dips in the Santa Susana Mountains and the southwestern edge of the east Ventura basin (Fig. 2). (The dip reversal south of the Pico anticline was a result of pre-Pleistocene south-directed blind thrusting; this structure has been tilted northeastward together with the Quaternary strata of the east Ventura basin¹³.) The upward-convex curvature of the Santa Susana fault marks the hinge of the fold.

Did the south-dipping blind thrust leave clues that could have identified it as a seismic source before the 1994 earthquake? Active north-dipping reverse faults in the western Transverse Ranges are characterized by uplift of only their hanging walls where they are not underlain by south-dipping blind thrusts.

FIG. 2 Cross-section through the western San Fernando Valley, Santa Susana Mountains and East Ventura basin (from refs 13 and 16). A-quality Northridge 1994 aftershocks from the Southern California Earthquake Center-Caltech catalogue lying within 1 km of the cross-section are plotted; location error bars are shown. Open triangles with vertical lines below show oil-well control, including dip of strata from dipmeter logs or cores.



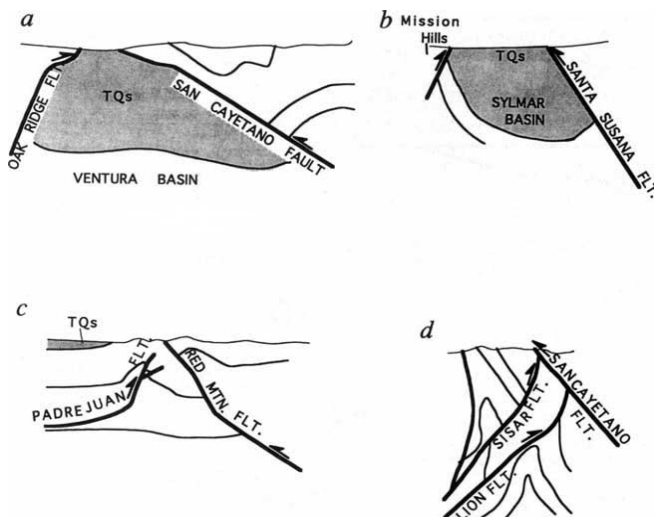


FIG. 3 Cross-sections at various scales (no vertical exaggeration) showing those north-dipping reverse faults that occur at the range front (a, b) and those that have uplifted footwalls because they are underlain by south-dipping blind thrusts (c, d). Compare with Santa Susana fault in Fig. 2. TQs, Pleistocene strata. Sources: a, b, (ref. 13); c, (ref. 17); d, (ref. 15).

Examples are the San Cayetano fault east of the Ojai Valley and east of Fillmore (Fig. 3a) and the Santa Susana fault north of the Sylmar basin (Fig. 3b). Where north-dipping faults are underlain by south-dipping blind thrusts, and both are active, both the hanging wall and footwall are uplifted. Examples are the Red Mountain fault underlain by the Padre Juan blind thrust (Fig. 3c)¹⁴ and the central San Cayetano fault underlain by the Searles-Lion fault system (Fig. 3d)¹⁵.

The footwall of the Santa Susana fault, raised during the 1994 earthquake¹, has been uplifted since the late Pleistocene along the entire length of the Santa Susana Mountains (Fig. 1), evidence of an underlying south-dipping blind thrust. (A north-dipping reverse fault beneath the Santa Susana fault might also uplift the footwall of the Santa Susana fault, but this possibility is unlikely, based on the presence of the south-dipping Frew fault along the entire length of the Santa Susana fault beneath the Santa Susana Mountains.) Footwall uplift, together with the thick Pliocene-Pleistocene sequences in the east Ventura basin (Fig. 2) and Sylmar basin, in contrast with the thin sequences of this age farther south, could have led investigators to a south-dipping source before the earthquake. However, it would not have led to a long-range forecast of a large earthquake along this fault for reasons outlined below.

The topographic expression of the south-dipping fault in the San Fernando Valley northeast of the epicentre is much less than it is in the Santa Susana Mountains. Indeed, the only topographic feature in the San Fernando Valley that could be attributed to uplift along the south-dipping fault is the Mission Hills south of the Sylmar basin, and this could equally well be related to displacement on the north-dipping Granada Hills-Mission Hills fault¹⁶. The slip rate on the Oak Ridge fault may decrease eastward from the Ventura basin, where the fault reaches the surface, to the Santa Susana Mountains and San Fernando Valley, where it does not. Slip rate on the south-dipping blind thrust beneath the Santa Susana Mountains is estimated as 1.4–1.7 mm yr⁻¹ based on a balanced cross-section modelling the east Ventura basin and Santa Susana Mountains as a fault-propagation fold controlled by the blind thrust⁴. This compares with 5 mm yr⁻¹ on the Oak Ridge fault in the Ventura basin. The strong gravity gradient that marks the fault dies out east of the Sylmar basin⁷, raising the possibility that the south-dipping blind thrust dies out also.

The Northridge earthquake, then, occurred in the least likely part of the Oak Ridge fault system because its long-term slip rate is about one-third of the rate farther west. East of the 1994 rupture, the south-dipping fault system may die out and not constitute a seismic source at all. However, there is a major danger of a rupture similar to that of 1994, on the Oak Ridge fault to the west in the Ventura basin, where the fault has a slip rate nearly three times higher than it is in the San Fernando Valley.

We have shown that footwall uplift characterizes three active north-dipping faults underlain by blind thrusts dipping in the opposite direction, and that other active faults not underlain by blind thrusts occur at mountain fronts and do not have uplifted footwalls. A search for footwall uplift elsewhere in California and in other regions with active reverse faults should reveal other blind thrusts similar to the seismic source of the Northridge earthquake. □

Received 17 February; accepted 13 December 1994.

1. Scientists of the U.S. Geological Survey and the Southern California Earthquake Center *Science* **266**, 389–397 (1994).
2. Yeats, R. S. *J. geophys. Res.* **93**, 12137–12149 (1988).
3. Yeats, R. S. *Nature* **366**, 299–301 (1993).
4. Davis, T. L. & Namson, J. S. *Nature* **372**, 167–169 (1994).
5. Donnellan, A., Hager, B. H. & King, R. W. *Nature* **366**, 333–336 (1993).
6. Donnellan, A., Hager, B. H., King, R. W. & Herring, T. A. *J. geophys. Res.* **98**, 21727–21739 (1993).
7. Langenheim, V. E., Griscom, A. & Jachens, R. C. 89th A. Mtg Seismol. Soc. Am., Program for Northridge Abstracts, abstr. 42 (1994).
8. Yeats, R. S. *US Geol. Surv. prof. Pap.* 1339, 137–160 (1987).
9. Yeats, R. S. in *Cenozoic Basin Development of Coastal California: Rubey Vol. VI* (eds Ingersoll, R. V. & Ernst, W. G.) 284–298 (Prentice-Hall, Englewood Cliffs, New Jersey, 1987).
10. Stein, R. S. & Yeats, R. S. *Scient. Am.* **260**(6), 48–57 (1989).
11. Okamura, Y., Satoh, M. & Miyazaki, J. *J. seismol. Soc. Japan 2nd Ser.* (in Japanese) **46**, 413–423 (1994).
12. Pfaffner, G. & Ward, S. N. *Tectonics* **11**, 709–718 (1992).
13. Yeats, R. S., Huftile, G. J. & Stitt, L. T. *Bull. Am. Ass. Petrol. Geol.* **78**, 1040–1074 (1994).
14. Yeats, R. S. & Grigsby, F. B. *Geol. Soc. Am. Centennial Field Guide* **1**, 219–223 (1987).
15. Huftile, G. J. *Bull. Am. Ass. Petrol. Geol.* **75**, 1353–1373 (1991).
16. Shields, K. E. thesis, Ohio Univ., Athens (1977).
17. Huftile, G. J. & Yeats, R. S. *J. geophys. Res.* (in the press).

ACKNOWLEDGEMENTS. We thank the petroleum industry for the data set that made this study possible. This research was supported by the Earthquake Hazards Reduction Program of the US Geological Survey and the Southern California Earthquake Center.

Elginerpeton pancheni and the earliest tetrapod clade

Per E. Ahlberg

Department of Palaeontology, Natural History Museum, Cromwell Road, London SW7 5BD, UK

THE first 'tetrapod-like' bones from the Scottish Upper Devonian site of Scat Craig (Scat Craig beds, Upper Frasnian) were figured and discussed in 1991 (ref. 1). Additional specimens have since been discovered that permit a formal description, and the background knowledge of Devonian tetrapod anatomy has greatly improved^{2–7}. These discoveries emphasize the uniqueness and phylogenetic importance of the Scat Craig material. *Elginerpeton pancheni* gen. et sp. nov., described here on the basis of cranial remains from Scat Craig, is, together with the fragmentary genus *Obruchevichthys* from the Upper Frasnian of Latvia and Russia^{1,8,9}, the oldest known stem tetrapod. *Elginerpeton* and *Obruchevichthys* form a clade that is the sister group of all other Tetrapoda. This contrasts with the later Devonian stem tetrapods, which all seem to represent separate plesions. *Elginerpeton* also has a unique, derived head morphology; it shows that the earliest phase of tetrapod evolution was accompanied by previously unrecognized morphological and phylogenetic diversification.

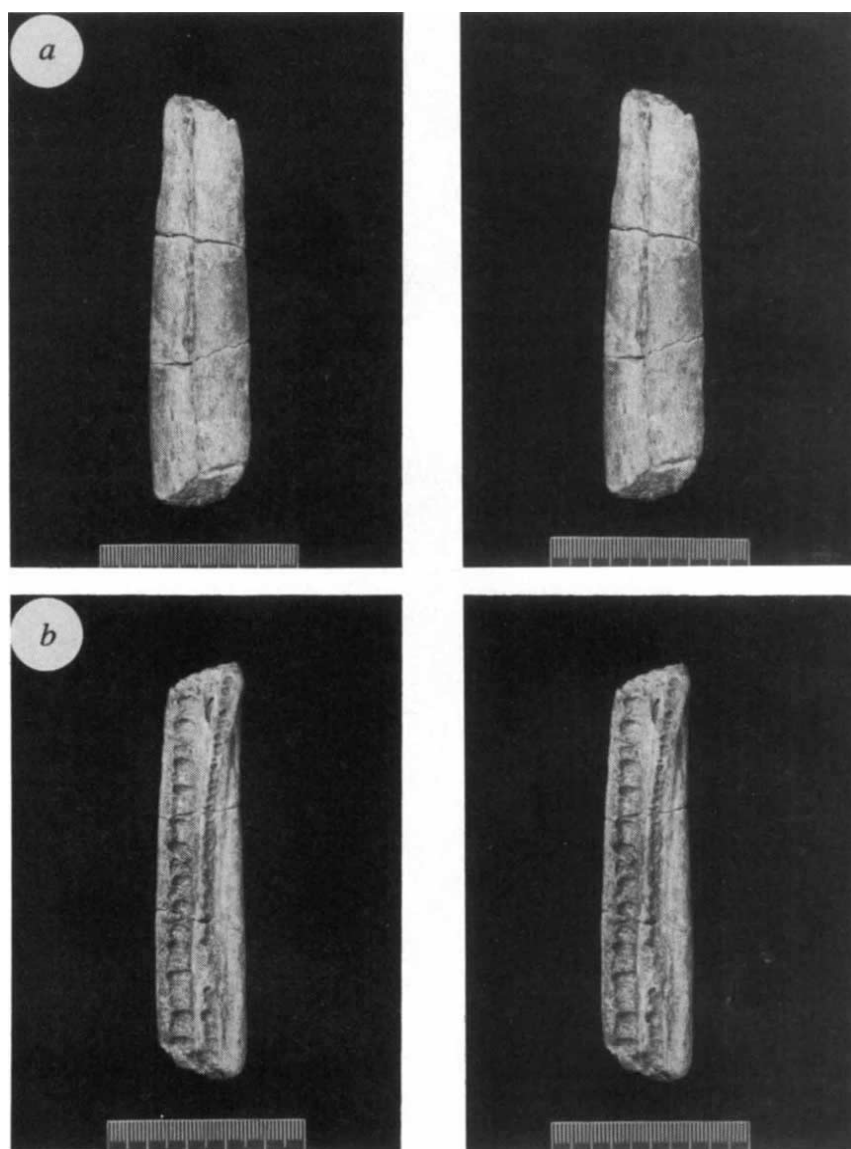


FIG. 1 *a, b*, RMS G 1967.17.1, holotype of *Elginerpeton pancheni*, an incomplete right mandible. Stereopair photographs in *a*, dorsal, and *b*, ventrolateral views. Scale bar, 50 mm.

Superclass Tetrapoda, Goodrich, 1930

Remark. The group is traditionally defined by the possession of limbs^{9,12}. However, there are also derived cranial tetrapod characters^{1,6}. *Elginerpeton* is assigned to the Tetrapoda on cranial characters, but associated tetrapod limb elements probably also belong to the genus.

Plesion Elginerpetontidae, nov.

Diagnosis. This plesion contains the genera *Elginerpeton* and *Obruchevichthys*. It is characterized by the following autapomorphies: the dentary and anterior infradentaries are swollen; the splenial and postsplenial are intimately fused; and a deep furrow marks the dentary-infradentary suture. A derived, narrow and dorsally placed prearticular is probably also characteristic for the Elginerpetontidae, but the condition in *Obruchevichthys* is not entirely clear (Fig. 2*e*). Unlike all other Tetrapoda, the Elginerpetontidae retain fangs on the anterior coronoid which are separate from the tooth row (Figs 2*f*, 4*a-c*).

Elginerpeton gen. nov.

Type species. *Elginerpeton pancheni* sp. nov.

Diagnosis. The genus can be recognized as a stem-group tetrapod by the possession of a fang pair and tooth row on the

parasymphysial plate; a pair of conspicuous foramina lateral and mesial to the posterior part of this plate; absence of exposed Meckelian bone dorsal to the prearticular; and irregular, radiating ornament on the premaxilla and infradentaries. None of these derived characters are found in non-tetrapod sarcopterygians^{1,6}, but they occur with minor variations in all known Devonian tetrapods^{6,13,14}. Most are secondarily lost in the crown group. *Elginerpeton* differs from all other known Devonian tetrapods in having rounded ventral infradentary margins and an accessory tooth row on the dentary. *Elginerpeton* and *Obruchevichthys* share several synapomorphies, which are defining characters of the Plesion Elginerpetontidae (see above).

Elginerpeton pancheni sp. nov.

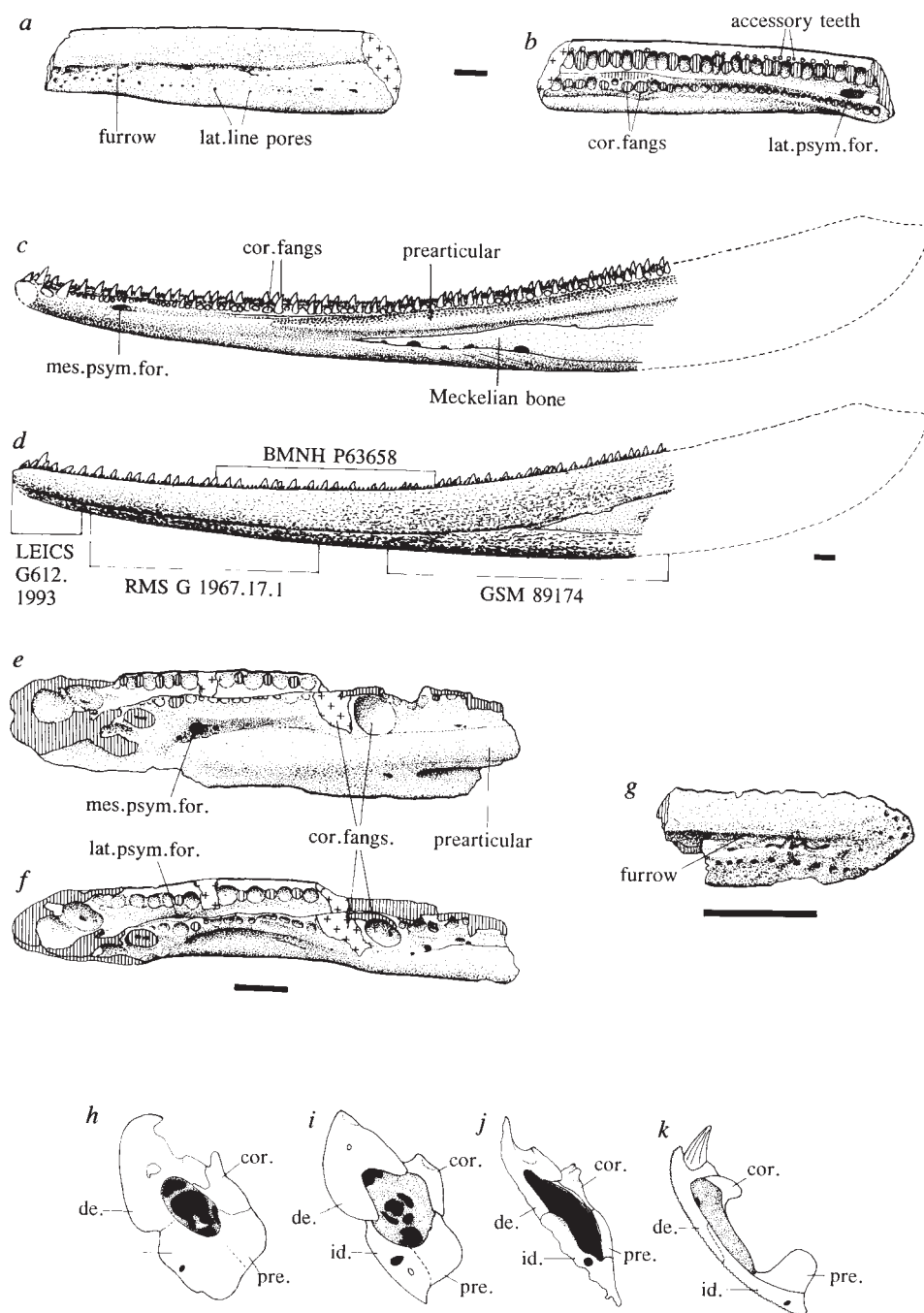
Polyplacodus Pander; Miles 1968, 8 (ref. 15).

'Tetrapod-like mandibular fragments'; Ahlberg 1991, 298 (ref. 1).

'Scat Craig tetrapod'; Ahlberg, Lukševičs and Lebedev 1994, 306 (ref. 6).

Etymology. Generic name after the town of Elgin and Greek *herpeton*, crawling. Specific name in honour of Alec L. Panchen, in recognition of his important contributions to the study of early tetrapods.

FIG. 2 a, b, RMS G 1967.17.1, holotype of *Elginerpeton pancheni*, in ventrolateral and dorsal views. Note furrow between dentary and infra-dentaries, lateral line pores (lat. line pores), small fang pair (cor. fangs) on the anterior coronoid, accessory tooth row on dentary, and large lateral parasymphysial foramen (lat.psym.for). c, d, reconstruction of the mandible of *Elginerpeton* in lateral and mesial views. Brackets indicate parts of the jaw preserved in different specimens. Note broad exposure of Meckelian bone ventral to the narrow prearticular. This is a derived pattern. e–g, *Obruchevichthys gracilis*. e, f, PIN 1491/51, holotype, Latvia, a partial mandible in mesial and dorsal views. The prearticular is narrow, and the coronoid tooth row loops round the fang pair (cor.fangs.). g, PIN 1491/52, St Petersburg District, anterior end of mandible, external view. All scale bars of a–g are 10 mm. Vertical hatching—broken bone. h–k, mandibular cross-sections of *Elginerpeton* (RMS G 1967.17.1), *Obruchevichthys* (PIN 1491/52), *Ichthyostega* (MGUH field no. 1618) and *Panderichthys* (ref. 22), all through anterior part of anterior coronoid. Not to scale. Stipple, Meckelian bone; black, hollow. *Elginerpeton* and *Obruchevichthys* share inflated dentary (de.) and infradentary (id.) bones. In *Panderichthys*, the coronoid (cor.) does not contact the prearticular (pre.).



Holotype. RMS G 1967.17.1, an incomplete left lower jaw ramus. Holotype housed at the National Museums of Scotland, Edinburgh.

Referred material. Table 1 lists the referred material, which includes upper and lower jaw elements.

Locality. Scat Craig is a streamside exposure on the Longmorn Burn near Fogwatt village south of Elgin, Morayshire, Scotland.

Horizon. Scat Craig Beds, Upper Frasnian, Upper Devonian¹⁵. The site yields psammosteids¹⁵, a vertebrate group that dies out at the Frasnian–Famennian boundary¹⁶. The age is ~368 Myr¹⁷.

Diagnosis. As for genus.

Jaw material of *Elginerpeton* can be recognized by its characteristic dentition. The most striking features of the lower jaw (Figs 1, 2a–d) are its slenderness, the grossly inflated infradentary bones (more extreme than in *Obruchevichthys*; Fig. 2h, i) and the broad exposure of Meckelian bone ventral to the prearticular (Fig. 2c). The overall jaw length can be estimated to 400 mm,

nearly twice that of *Ventastega* or *Ichthyostega* (Fig. 4d–h). These features are unparalleled in Famennian tetrapods. In large specimens of *Elginerpeton*, the coronoid fangs are scarcely larger than the marginal teeth. The small jaw BMNH P62809 resembles *Obruchevichthys* in having proportionately larger fangs¹. This may suggest negative allometry during growth, but as the specimen lacks some diagnostic regions its attribution to *Elginerpeton* is slightly uncertain.

Elginerpeton and *Obruchevichthys* both have a coronoid fang pair that is separate from the tooth row; this is a primitive feature unknown in other tetrapods^{3,6,13,14} (Figs 1, 2f, 4a–c). A probable vomer of *Elginerpeton* (GSM 89132) shows a corresponding pattern, again uniquely primitive within the Tetrapoda. This strongly suggests that *Elginerpeton* and *Obruchevichthys* are the joint sistergroup to all other tetrapods (Fig. 4i).

The premaxilla (Fig. 3) differs in shape from those of other basal tetrapods. In *Ichthyostega*¹⁴, *Acanthostega*⁴, *Ventastega*⁶

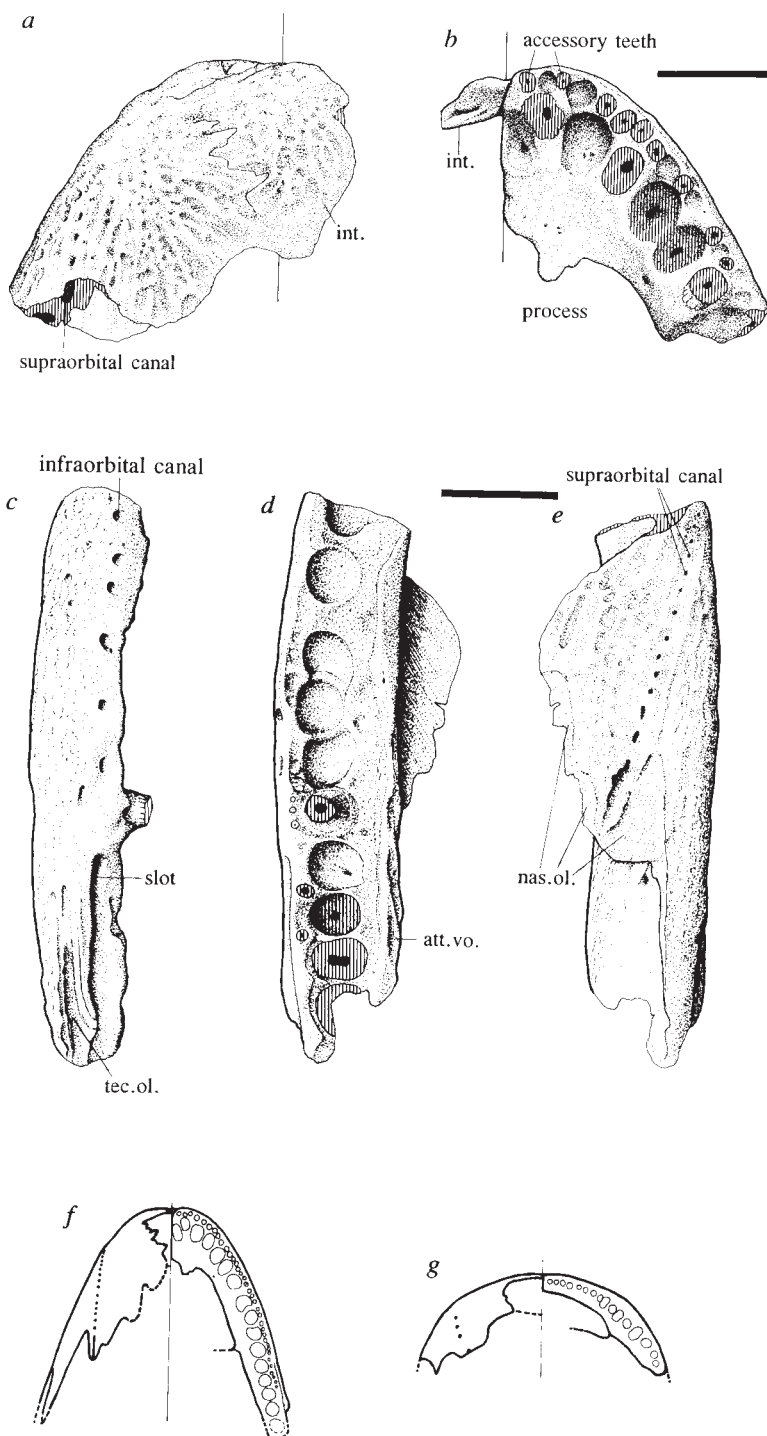


FIG. 3 The premaxilla of *Elginerpeton*. *a*, *b*, OUM Geol. Coll. D796, anterior part of premaxilla in dorsal and ventral views. Note the radiating dermal ornament, characteristic of early tetrapods. There is a single internasal bone (int.) as in *Ichthyostega*¹⁴, whereas *Acanthostega* has two⁴. The accessory tooth row, which matches that on the dentary, is unknown in other tetrapods. The posterior process on the ventral surface may have sutured with an anterior extension of the vomer as in *Acanthostega*⁴. *c*–*e*, BMNH P9776, posterior part of premaxilla in lateral, ventral and dorsal views. The overlap facets for the tectal (tec.ol.) and nasal (nas.ol.) bones correspond positionally to those of *Ventastega*⁶, but are much more elongate. It is not clear whether a lateral rostral was present. Note the long open slot for the posterior part of the infraorbital canal. There is a large attachment area for the vomer (att.vo.). Scale bars for *a*–*e* are 10 mm. Thick outline, true edge; vertical hatching, broken bone. *f*, *g*, Reconstructions in dorsal (left half) and ventral (right half) view of the snouts of *Elginerpeton* and *Ventastega*. The proportions of the two are very different. Not to scale.

and *Tulerpeton*³, the snout is broad and spade-shaped (Fig. 4g). *Elginerpeton* has a pointed snout and the posterior part of the premaxilla is very elongated (Fig. 3a–f). The combined evidence from the snout and mandible indicates a narrow triangular skull shape (Fig. 4g).

Although the postcranial tetrapod elements from Scat Craig cannot formally be attributed to *Elginerpeton*, they are the right size (that is, very large) and most probably pertain to it. They include an *Ichthyostega*-like tibia¹, robust ilia, a neural arch with weak zygapophyses, and incomplete shoulder girdles comparable to *Hynerpeton*⁷. All indicate a postcranium broadly comparable with those of *Ichthyostega*^{14,18} and *Acanthostega*^{2,18,19}; only a humerus¹ and a femur appear significantly more primitive. This agrees well with the cranial material. Scat Craig has yielded

no remains of hands or feet, but Frasnian Australian trackways show digit impressions²⁰.

Eight Devonian tetrapods are now known: *Elginerpeton* and *Obruchevichthys* from the Upper Frasnian, *Metaxygnathus* and *Hynerpeton* from the Lower to Middle Famennian^{7,13}, and *Ventastega*, *Acanthostega*, *Ichthyostega* and *Tulerpeton* from the Upper Famennian^{2–6,9,14,19,21}. All except perhaps *Tulerpeton* (M. I. Coates, personal communication) are members of the tetrapod stem group. The peculiar specializations of *Elginerpeton*, and its apparent sistergroup relationship with *Obruchevichthys*, cast an interesting new light on the earliest phase of tetrapod evolution.

The Famennian tetrapods seem to form a wholly paraphyletic array. *Tulerpeton*, *Ichthyostega* and *Acanthostega* clearly represent separate plesions⁹, while *Ventastega* and *Metaxygnathus*

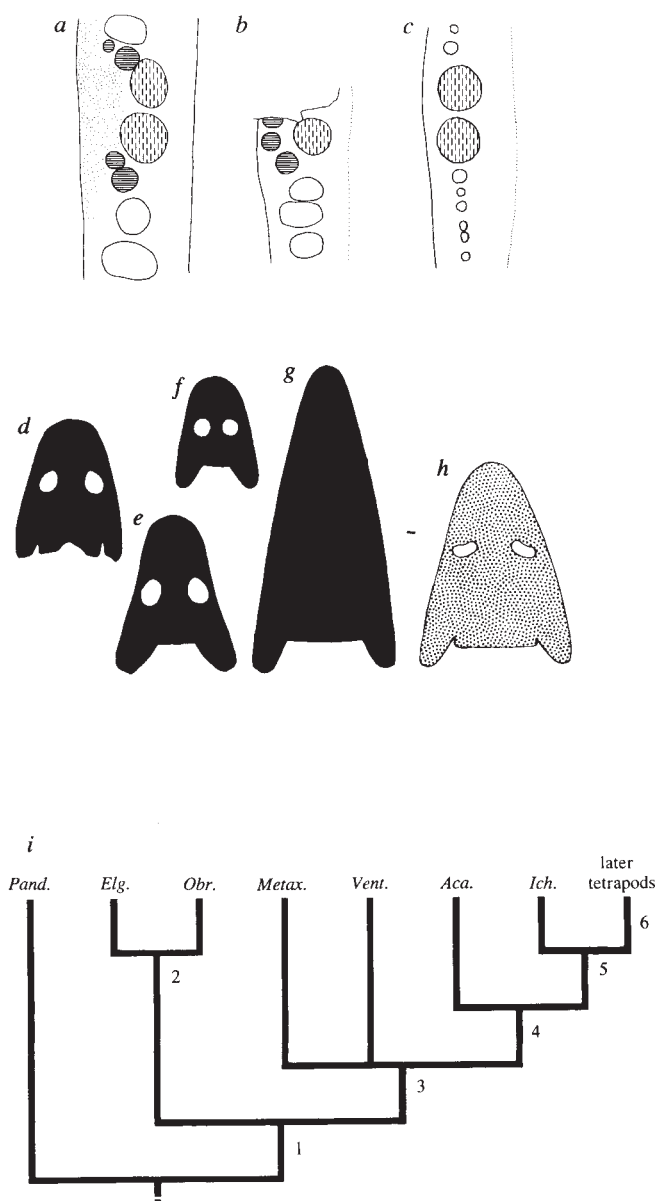


FIG. 4 a–c, Coronoid dentitions of *Elginerpeton* (a, b) and *Ventastega* (c). Lateral margin to the left (not to scale). In *Elginerpeton*, the coronoid tooth row continues lateral to the fangs on the anterior coronoid. (Fangs, vertical dashes; teeth lateral to fangs, horizontal hatching; damaged area, dot shading.) In *Ventastega* the fangs (vertical hatching) lie in the tooth row. d–h, Reconstructed skull outlines of d, *Ichthyostega* (from ref. 14); e, *Ventastega* (from ref. 6); f, *Acanthostega* (from ref. 4); g, *Elginerpeton*; and h, *Panderichthys* (from ref. 8). The orbital position is unknown in *Elginerpeton*, and the posterior margin of the skull is conjectural. The extrascapulars and opercular series of *Panderichthys* are not shown. Scale bar, 10 mm. i, Tentative cladogram of *Panderichthys*, *Elginerpeton*, *Obruchevichthys*, *Metaxygnathus*, *Ventastega*, *Acanthostega*, *Ichthyostega* and higher tetrapods. The characters supporting the nodes are 1, tooth row on parasymphysial plate, parasymphysial foramina, no exposed Meckelian bone dorsal to prearticular, radiating irregular dermal ornament, large nasal bones; 2, inflated dentary and infradentary bones, furrow along suture between them; 3, coronoid fangs in tooth row, broad spade-shaped snout; 4, coronoid fangs lost; 5, olecranon process on ulna, radius not longer than ulna; 6, scapular blade, six or fewer digits. As the genera between nodes 1 and 4 lack articulated postcrania, the exact sequence of postcranial character acquisition is uncertain. But if the postcranial tetrapod material from Scat Craig belongs to *Elginerpeton* (as seems likely), the following characters can be added at node 1: sacrum, laterally facing acetabulum, femur with adductor blade and intercondylar fossa, tarsal bones, plate-like scapulocoracoid with subscapular fossa, zygapophyses.

TABLE 1 Specimens attributed to *Elginerpeton pancheni*

Specimen number	Institution	Elements(s)
RMS G 1967.17.1	National Museums of Scotland, Edinburgh	Incomplete left mandible
RMS G 1891.92.444	National Museums of Scotland, Edinburgh	Incomplete right premaxilla
RMS G 1994.139.2	National Museums of Scotland, Edinburgh	Coronoid fragment
BMNH P8277	Natural History Museum, London	Twelve mandibular fragments
BMNH P9776	Natural History Museum, London	Incomplete right premaxilla
BMNH P62809	Natural History Museum, London	Small incomplete right mandible
BMNH P63658	Natural History Museum, London	Incomplete right mandible
BMNH P8267	Natural History Museum, London	Jaw fragment, probably maxilla
GSM 89132	British Geological Survey, Nottingham	Palatal bone, probably vomer
GSM 89174	British Geological Survey, Nottingham	Incomplete right mandible
GSM 89175	British Geological Survey, Nottingham	Fragment of right mandible
OUM Geol. Coll. D796	University Museum, Oxford	Incomplete left premaxilla and internasal
LEICS G612.1993	Leicestershire Museums, Leicester	Left dentary fragment

may fall below *Acanthostega* on the tetrapod stem⁶. Each genus is known from only one location. This suggests the existence of a clear stem lineage with short-lived, geographically restricted branches. The genera display considerable postcranial variation, but their jaw morphology and head shape are strikingly uniform^{3–5,14} (Fig. 4d–f); these common features must have characterized the actual stem lineage, and may be evidence for a conserved mode of prey capture.

The Frasnian picture can now be seen to be quite different. All the known Frasnian tetrapod body fossils, ranging from Scotland to western Russia, belong to the elginerpetontids *Elginerpeton* and *Obruchevichthys*. Such wide dispersal of a tetrapod 'side branch' is not repeated in the fossil record until the Visean⁹. Furthermore, the head shape, mandibular construction and size of *Elginerpeton* are all different from those of Famennian tetrapods. Outgroup comparison with panderichthyids^{8,22} (Fig. 4h) and osteolepiforms¹⁴ suggests that both tetrapod morphologies are derived, but the *Elginerpeton* pattern is the more profoundly modified. The elginerpetontids thus seem to be strongly divergent from the tetrapod stem: this was probably reflected by their mode of life.

Although the elginerpetontids are the sistergroup of all other tetrapods, only one character supports the internode between them and *Ventastega* or *Metaxygnathus* (Fig. 4i). The postcranial tetrapod material from Scat Craig is likewise almost as advanced as that of *Acanthostega* or *Ichthyostega*. This suggests that *Ventastega*, *Acanthostega* and *Ichthyostega* are conservative survivors from a Lower Famennian (or earlier) radiation. The narrow time span and wide morphological gap between the elginerpetontids and the Lower Frasnian panderichthyid 'fishes', the next lower plesion on the tetrapod stem⁹, suggests either a very rapid transformation of the postcranial skeleton during the Lower–Middle Frasnian, or a pre-Frasnian date for the panderichthyid–tetrapod split.

The Scat Craig fossils show that tetrapod evolution was well under way before the end of the Frasnian. Two groups, at least, were in existence: the stem lineage to Famennian and later tetrapods, and the widespread and morphologically divergent elginerpetontids. The recognition of the Elginerpetontidae provides the first real evidence of biogeographic and biostratigraphic patterns at the beginning of tetrapod evolution. The characteristics, range and eventual fate of this clade should be a key target for future research. □

Received 27 October; accepted 5 December 1994.

- Ahlberg, P. E. *Nature* **354**, 298–301 (1991).
- Coates, M. I. & Clack, J. A. *Nature* **352**, 234–235 (1991).
- Lebedev, O. A. & Clack, J. A. *Palaeontology* **36**, 721–734 (1993).
- Clack, J. A. *Meddr. Grønland Geoscience* **31**, 3–24 (1994).
- Clack, J. A. *Nature* **369**, 392–394 (1994).
- Ahlberg, P. E., Lukševičs, E. & Lebedev, O. *Phil. Trans. R. Soc. B* **343**, 303–328 (1994).
- Daeschler, E. B., Shubin, N. H., Thomson, K. S. T. & Amaral, W. W. *Science* **265**, 639–642 (1994).
- Vorobyeva, E. I. *Trudy paleont. Inst.* **163**, 1–239 (1977).
- Ahlberg, P. E. & Milner, A. R. *Nature* **368**, 507–514 (1994).
- Gaffney, E. S. *Bull. Carnegie Mus. nat. Hist.* **13**, 92–105 (1979).
- Panchen, A. L. & Smithson, T. R. *Biol. Rev.* **62**, 341–438 (1987).
- Godfrey, S. J. *Phil. Trans. R. Soc. B* **323**, 75–133 (1989).
- Campbell, K. S. W. & Bell, M. W. *Alcheringa* **1**, 369–381 (1977).
- Jarvik, E. *Basic Structure and Evolution of Vertebrates Vol. 1* (Academic, London, 1980).
- Miles, R. S. *The Old Red Sandstone Antiarchs of Scotland: Family Bothriolepididae* (Palaeontographical Society, London, 1968).
- Gross, W. *Abh. dt. Akad. Wiss. Math.-naturw. Kl.* **1**, 3–130 (1950).
- Harland, W. B. et al. *A Geologic Time Scale 1989* (Cambridge University Press, 1989).
- Coates, M. I. & Clack, J. A. *Nature* **347**, 66–69 (1990).
- Coates, M. I. in *Developmental Pattern in the Vertebrate Limb* (eds Hinchliffe, J. R. et al.) 325–337 (Plenum, New York, 1991).
- Warren, J. W. & Wakefield, N. A. *Nature* **238**, 469–470 (1972).
- Lebedev, O. A. *Dokl. Akad. Nauk. SSSR* **278**, 1407–1473 (1984).
- Gross, W. *Abh. preuss. Akad. Wiss. Math.-naturw. Kl.* **1941**, 1–51 (1941).

ACKNOWLEDGEMENTS. I thank M. Taylor (National Museums of Scotland) and S. Tunnicliff (British Geological Survey) for lending me *Scat Craig* specimens, J. Clack and M. Coates for access to undescribed *Acanthostega* material, E. Jarvik (Swedish Museum of Natural History) for access to *Ichthyostega*, S. Finney for preparation work, R. Hitchin for assistance with illustrations, and P. Crabb (Natural History Museum) for taking the photographs. I also want to thank the National Museums of Scotland for undertaking a major excavation programme at *Scat Craig*.

Attraction of kestrels to vole scent marks visible in ultraviolet light

Jussi Viltala*, Erkki Korpimäki††, Päivi Palokangas† & Minna Kolvula†

* Konnevesi Research Station, Department of Bio- and Environmental Sciences, University of Jyväskylä, FIN-44300 Konnevesi, Finland
† Laboratory of Ecological Zoology, Department of Biology, University of Turku, FIN-20500 Turku, Finland

IN northern Europe, broad four-year oscillations in small rodent and raptor populations are synchronous over hundreds of square kilometres^{1–6}. Crashes in vole populations can induce wide emigration (>1,000 km) of their predators^{7–9}, but almost nothing is known about how predators rapidly detect areas of vole abundance. Here we report on laboratory and field experiments on voles (*Microtus agrestis*) and kestrels (*Falco tinnunculus*). Voles mark their runways with urine and faeces, which are visible in ultraviolet light. Wild kestrels brought into captivity were able to detect vole scent marks in ultraviolet light but not in visible light. In the field, kestrels hunted preferentially near experimental nest-boxes where artificial trails were treated with vole urine and faeces. We suggest that kestrels flying over an area can see and use vole scent marks to assess vole numbers. This ability would enable kestrels to 'screen' large areas in a relatively short time. Our results provide a novel explanation for how raptors detect patches of high vole densities without prior knowledge of local food resources.

The Eurasian kestrel (*Falco tinnunculus*) (hereafter kestrel) is a widespread, open-country raptor that is migratory in Fennoscandia. Kestrels detect prey visually, feeding primarily on small mammals that they capture on the ground¹⁰. Their main hunting modes are wind-hovering, soaring and perching¹⁰. Small rodents mark their runways with urine and faeces^{11,12}. These marks are visible in ultraviolet light (wavelengths of 320–400 nm)¹³ (Fig. 1). The above-ground runways of voles are clearly visible to

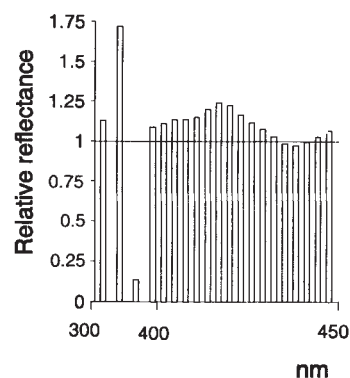


FIG. 1 Reflection of scent marks of a male field vole on cardboard used in a laboratory experiment, relative to clean cardboard. A value of 1 represents clean cardboard (horizontal line), values <1 denote greater absorption and those >1 greater reflection than clean cardboard. Each column is a mean of five measurements (Spectroradiometer SR-500). The bandwidth in visible light was 3 nm but the UV filters penetrated a 10-nm band. At visible wavelengths, the reflection remained close to 1 up to 690 nm. The figure shows that there are contrasts in UV wavelengths that facilitate the detection of vole scent marks by a UV-sensitive spectator. There is a clear absorption at 370 nm (in the UV wavelength range of the laboratory experiment). The clean cardboard also reflected UV light considerably, so the reflection peak near 340 nm may also be visible in the wild in UV-absorbing green vegetation.

raptors, especially in spring. Their bottoms (3–4 cm wide) are soaked by urine and faeces. We tested the hypothesis that kestrels use vole scent marks visible in ultraviolet light as a cue for discovering prey patches in laboratory and field experiments.

For the laboratory investigation we used a room measuring 5 × 4 × 2.5 m at the Konnevesi Research Station, with the walls covered by black paper to exclude other visible cues for orientation. There was no natural light in the room. Kestrels were observed from a dark hide through a 30 × 30-cm wire-mesh window with 1 mm openings so that the observer was invisible. This experimental situation was not unnatural, as wild kestrels sometimes hunt inside small barns.

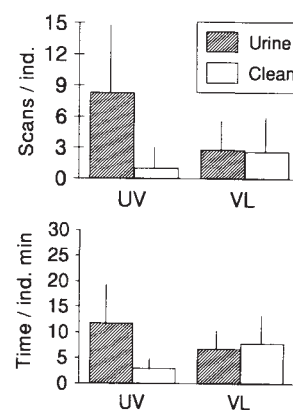
In the laboratory experiment, captured wild kestrels were each given a choice of four adjacent identical arenas: arena in ultraviolet light with vole trails, clean arena in ultraviolet light, vole trail arena in visible light, and clean arena in visible light (Fig. 2). In terms of both the time spent above the arenas and the number of scans, kestrels preferred the vole-trail arena illuminated by ultraviolet light over other arenas, whereas the clean arena illuminated by ultraviolet light was less favoured than any other (Fig. 2); there was no difference in preference between the two arenas illuminated by visible light. We propose that kestrels found the vole-trail arena in ultraviolet light more interesting than the clean arena under the same lighting conditions because they could see fresh signs of voles in the former. Their interest in the two arenas in visible light was intermediate, possibly because they could not discern whether these arenas contained signs of voles.

The field experiment was conducted in the Alajoki study area (63° N, 23° E), which is a large (100 km²) tract of level farmland in western Finland. Because most fields are ploughed in autumn, the only habitat suitable for voles during winter and early spring is along ditches¹⁴. Kestrels can breed in nest-boxes fastened on barns and individual trees⁵. They are highly nomadic, as shown by a mean turnover of 75% for males and 92% for females during the breeding seasons of 1983 to 1992 (ref. 15). Annual breeding densities of kestrels at Alajoki vary from 4 to 98 nests per 100 km² and synchronously with vole densities^{5,6}.

In spring 1993, vole densities at Alajoki were extremely low, so no vole trails (runways) were evident. In early April (before the arrival of kestrels), we chose 45 kestrel nest-boxes (in an area of 15 km²) and randomly placed them into one of three

† To whom correspondence should be addressed at the Department of Zoology, University of British Columbia, 6270 University Boulevard, Vancouver, BC V6T 1Z4, Canada.

FIG. 2 Mean time (min, +s.d.) spent by the 19 kestrels above four adjacent 1 × 2 m arenas (below) and mean number of scans (+s.d.) of the falcons on different arenas (top). The four options were: dry vole trails (urine and faeces) in UV light, clean (no vole trails) arena in UV light, dry vole trails (urine and faeces) in visible light (VL) and clean arena in visible light (VL). Vole trail arenas had been occupied by seven field voles for 17 hours before the experiment and two had been without voles. The voles were removed before introducing kestrels to the arenas. Two 160-W black-light UV bulbs (Philips MLW 160W; spectral energy distribution shown in Fig. 4a) illuminated one arena of each group. The intensity of UV (320–400 nm) on the arena floor below the lamps varied from 0.016 to 0.025 mW cm⁻². Two ordinary 60-W bulbs (OSRAM 9W3, 60W; irradiance spectrum shown in Fig. 4b) emitting <1% UV illuminated the other arenas. The sockets of the lamps were 115 cm above the bottom of the arenas. The intensity of VL (400–700 nm) on the arena floor below the lamps varied from 0.55 to 0.67 mW cm⁻². The temperature difference on the bottom of the arenas between UV and VL treatments was <0.2 °C. On UV arenas there also was some VL scattered from the lamps above the VL arenas. It was enough to mask possible fluorescence and to facilitate recording of falcon behaviour. There was also UV scattering on VL arenas, but our gauge was not sensitive enough to measure it. Each arena had a hard brown cardboard floor and identical nest-box for voles. Arena positions were changed after each exposure. There was a perch for kestrels above each arena. Nineteen kestrels (8 young of the year (age 6–8 weeks) and 11 adults), starved on the day before the experiment, were individually introduced to the experimental room twice for 15 min each time. The time between introductions of a falcon was at least 48 h. The time spent above each arena was recorded, as was the number of scans by each falcon; that is, the times when the kestrel pointed its eyes to an arena, bobbing its head to estimate the distance to the target, an easily discernible hunting mode of the kestrel¹⁰. The means of the distributions were compared

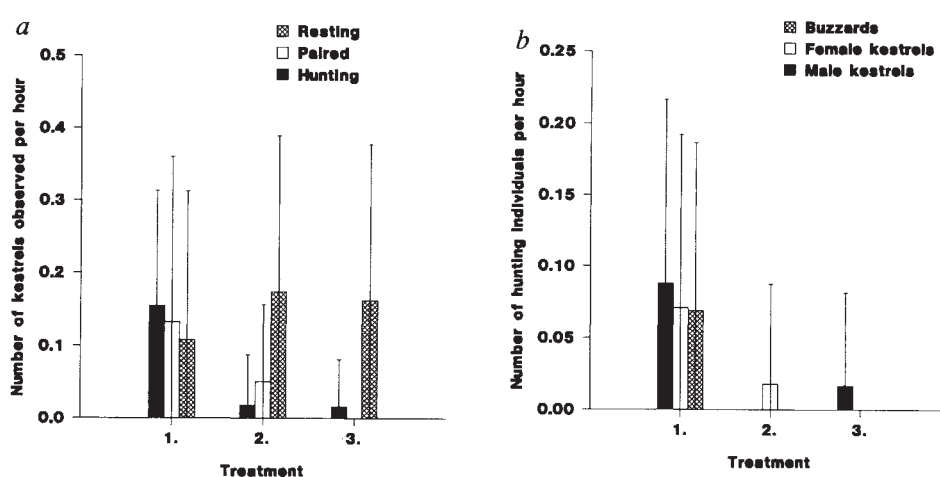


after square-root transformation by SPSS/two-way ANOVA for repeated measures (d.f. = 1, 18). For time spent above the arenas, the effect of light was nearly significant ($F = 3.13$, 2-tailed $P = 0.095$) and the effect of vole scent marks highly significant ($F = 32.04$, $P < 0.001$; interaction term $F = 15.95$, $P = 0.001$). For the number of scans, the effect of light was not significant ($F = 0.24$, $P = 0.63$), whereas the effect of scent marks was significant ($F = 10.42$, $P = 0.005$; interaction term $F = 18.09$, $P = 0.001$). The contrasted differences between arenas in UV light with and without scent marks were: $F = 31.86$, $P < 0.001$ for time spent and $F = 44.44$, $P < 0.001$ for the number of scans, and between arenas in VL with and without scent marks: $F = 0.56$, $P = 0.47$ for time spent and $F = 0.20$, $P = 0.66$ for the numbers of scans.

experimental categories: (1) artificial vole trails with urine and faeces; (2) artificial vole trails without urine and faeces; and (3) no vole trails (control) (Fig. 3a). We counted raptors near the boxes during 24 mornings using binoculars and a telescope. Each nest-box was observed for 15–30 min each morning for a total of 6–7 hours. During observation periods, we recorded the species, sex and behaviour (hunting (hovering, soaring and perched-hunting), paired (both sexes near box) or resting) of all

raptors seen near the boxes. It is fairly obvious when a kestrel is hunting from a perch, because it has an upright posture, continually bobs its head and scans the ground, and often alters perch¹⁰. All these behaviours are different from those of resting kestrels, which sit hunched up with feathers fluffed out, or preen. Because the study area consisted mostly of open agricultural fields (73%), we detected most (if not all) of the raptors present during observation periods.

FIG. 3 a, Log-transformed ($\log(n+1)$) number of hunting, paired and resting kestrels observed per hour near nest-boxes. Boxes were randomly assigned to one of three experimental treatments (15 boxes per treatment), and the bars show the mean (+s.d.) number of kestrels per box. In treatment 1, we created four artificial vole runaways in ditches within 100 m of each nest-box by cutting grass on four 10-m long and 4–6-cm wide trails. Trails were treated with straw taken from cages that had contained field voles held over the previous winter at the Konnevesi Research Station, so the straw was saturated with vole urine and faeces. In treatment 2, artificial vole trails were created at each nest-box as in treatment 1, except trails were treated with straw taken from surrounding fields and wetted with water from nearby ditches. Water blended with vole urine and faeces (treatment 1) or taken from ditches (treatment 2) was added to the trails once a week over 5 weeks. Nest-boxes assigned to treatment 3 served as controls, for which no vole trails, or urine and faeces or water treatments were made. ANOVA indicated that the log-transformed number of hunting ($F_{2,42} = 8.36$, 2-tailed $P = 0.001$) and paired ($F_{2,42} = 3.23$, $P = 0.049$) kestrels differed among treatments, whereas the number of resting kestrels was equal among treatments ($F_{2,42} = 0.41$, $P = 0.67$). In general, hunting kestrels came preferentially near boxes containing artificial trails treated with vole urine and faeces (Tukey test for the difference between treatments 1 and 2, 2-tailed $P = 0.003$; between treatments 1 and 3, $P = 0.003$). b, Log-transformed ($\log(n+1)$) mean (+s.d.) number of hunting male kestrels, female



kestrels and rough-legged buzzards observed per hour near nest-boxes, classified by experimental treatments as outlined in a (1 represents artificial vole trails treated with straw and vole urine and faeces; 2, artificial vole trails treated with straw and water; and 3, control). The log-transformed number of male kestrels (ANOVA, $F_{2,42} = 4.72$, $P = 0.014$) and the number of rough-legged buzzards (Kruskal–Wallis test, $H = 8.56$, $P = 0.014$) differed significantly among treatments, whereas the log-transformed number of female kestrels approached significance ($F_{2,42} = 3.11$, $P = 0.055$). Male kestrels hunted more frequently near nest-boxes that had artificial trails treated with vole urine and faeces (Tukey-test for the difference between the treatments 1 and 2, 2-tailed $P = 0.016$; between the treatments 1 and 3, $P = 0.06$).

Kestrels were observed near 27 experimental nest-boxes. They hunted preferentially near nest-boxes where artificial trails had been created using vole urine and faeces (Fig. 3a). In contrast, hunting kestrels largely avoided areas near boxes with trails but no urine and faeces, or with no trails (Fig. 3a). Thus, kestrels were not using the trails themselves as cues for hunting. Paired kestrels also tended to occupy boxes near trails treated with vole urine and faeces, whereas resting kestrels chose boxes irrespective of treatment (Fig. 3a). Male kestrels provide for their mates from pair formation onwards, during which time females do not hunt but remain near their nest-box in preparation for egg-laying^{10,16}. Therefore, the number of male kestrels hunting near boxes was greater than the number of females. This may explain why treatments differed in the number of hunting kestrels, significantly for males (which preferred to hunt near urine- and faeces-treated trails) but only marginally for females (Fig. 3b). In addition, the four rough-legged buzzards (*Buteo lagopus*) seen hunting near nest-boxes were all near sites treated with vole urine and faeces (Fig. 3b).

We have provided the first experimental evidence, to our knowledge, of a wild raptor using vole trail marks to select hunting patches and potential nest-sites. We suggest that scent marks of voles act as visible cues to kestrels, especially in spring when these are not covered by grass. In the laboratory experiment, we did not detect any peak of visible light in the ultraviolet arenas that might explain our results (Fig. 4). It is known that mouse urine fluoresces in blue¹³, and we have found a dim blue fluorescence in the urine and faeces of voles. It is possible that the ability of kestrels to detect vole scent marks may not depend entirely on ultraviolet vision. The faint fluorescence is just visible to a human spectator when ultraviolet lamps are the only source of light: light from ordinary 60 W bulbs masks it completely. At Alajoki during our field experiment, the visible sunlight was so strong that any fluorescence was totally masked; we therefore consider it unlikely that fluorescence was the reason for our results. An alternative explanation, that kestrels detect vole scent marks by olfaction, is also unlikely, as in our laboratory experiment kestrels were able to distinguish vole scent marks in ultraviolet but not visible light. The biological significance of ultraviolet vision in higher vertebrates is poorly understood: it may play a role in orientation^{17,18}, food detection^{19,20} and intraspecies

communication^{21,22} (reviewed in refs 23, 24). The absorbance or reflection of ultraviolet light by a food item could make it more conspicuous to a consumer. Although the eye structure of diurnal raptors (order Falconiformes) has not been tested systematically for sensitivity to ultraviolet light, most other diurnal birds studied so far have proved to be sensitive²⁴⁻²⁶. We therefore propose that, in the presence of ultraviolet light, diurnal raptors can easily see areas stained with fresh vole urine and faeces, and that they use these marks as visual cues when searching for areas of vole abundance. This ability would enable raptors to evaluate large areas in a relatively short time and would explain how nomadic raptors find patches of high vole abundance without prior knowledge of local food conditions. It has been suggested that scent marks of other vertebrates may also be visible in ultraviolet light²⁷, so this kind of hunting-site detection may not be uncommon. □

Received 10 June; accepted 7 December 1994.

1. Finerty, J. P. *The Population Ecology of Cycles in Small Mammals. Mathematical Theory and Biological Fact* (Yale University Press, New Haven and London, 1980).
2. Kalela, O. *Ann. Acad. Sci. Fenn. A IV* **66**, 1-38 (1962).
3. Hanski, I. *et al. Nature* **364**, 232-235 (1993).
4. Korpimäki, E. *Oikos* **45**, 281-284 (1985).
5. Korpimäki, E. & Norrdahl, K. *Ecology* **72**, 814-826 (1991).
6. Korpimäki, E. *J. Anim. Ecol.* **63**, 619-628 (1994).
7. Andersson, M. *J. Anim. Ecol.* **49**, 175-184 (1980).
8. Ward, R. M. P. & Krebs, C. J. *Can. J. Zool.* **63**, 2817-2824 (1986).
9. Korpimäki, E., Lagerström, M. & Saurola, P. *Ornis Scand.* **18**, 1-4 (1987).
10. Village, A. *The Kestrel* (Poyser, London, 1990).
11. Johnson, R. P. *Behaviour* **55**, 81-93 (1975).
12. Rozenfeld, F. M., Le Boulange, E. & Rasmont, R. *Can. J. Zool.* **65**, 2594-2601 (1987).
13. Desjardins, D., Maruniak, J. A. & Bronson, F. H. *Science* **182**, 939-941 (1973).
14. Norrdahl, K. & Korpimäki, E. *Oikos* **67**, 149-158 (1993).
15. Korpimäki, E. *Oecologia* **77**, 278-285 (1988).
16. Palokangas, P., Alatalo, R. V. & Korpimäki, E. *Anim. Behav.* **43**, 659-665 (1992).
17. Kreithen, M. L. & Eisner, T. *Nature* **272**, 347-348 (1978).
18. Hawryshyn, C. W. *et al. J. comp. Physiol. A* **166**, 565-574 (1990).
19. Goldsmith, T. H. *Science* **207**, 786-788 (1980).
20. Burkhardt, D. *Naturwissenschaften* **69**, 153-157 (1982).
21. Burkhardt, D. *J. comp. Physiol. A* **164**, 787-796 (1989).
22. Fleishman, L. J., Loew, E. R. & Leal, M. *Nature* **365**, 397 (1993).
23. Jacobs, G. H. *Am. Zool.* **32**, 544-554 (1992).
24. Bennett, A. T. D. & Cuthill, I. C. *Vision Res.* **34**, 1471-1478 (1994).
25. Chen, D.-M., Collins, J. S. & Goldsmith, T. H. *Science* **225**, 337-340 (1984).
26. Maier, E. J. & Bowmaker, J. K. *J. comp. Physiol. A* **172**, 295-301 (1993).
27. Alberts, A. C. *Anim. Behav.* **38**, 129-137 (1989).

ACKNOWLEDGEMENTS. We thank J. Kemppi, P. Niutanen and A. Sirkka for help in the laboratory; K. Lahti, E. Rautiainen, J. Ryssy and P. Tolonen for assistance in the field work; and R. V. Alatalo, A. T. D. Bennett, I. C. Cuthill, S. Daan, E. Haukioja, T. Horne, R. A. Ims, G. H. Jacobs, J. Jokela, H. Karppinen, J. S. Marks, S. Neuvonen, H. Pietiäinen, J. Pusenius, J. v. Weissenberg and H. Ylönen for invaluable comments on the manuscript. The field experiment was financially supported by the Academy of Finland (E.K.). Kestrels were captured and kept in captivity with the permission of the Finnish Ministry of Environment.

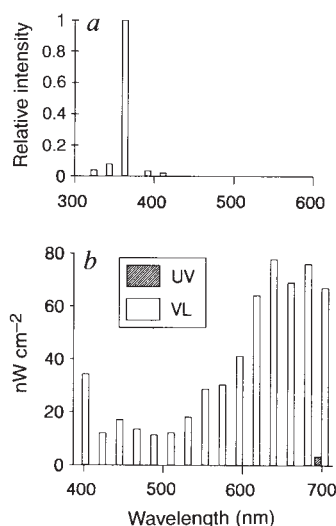


FIG. 4 a, Relative spectral energy distribution of the UV lamp (Philips MLW 160 W, 220-240 V) according to the manufacturer. b, The irradiance spectra between 400 and 700 nm on the bottom of UV arenas (UV) and visible light (VL) arenas measured with a series of interference filters (each penetrating ~15-nm band) immediately below the UV and VL lamps, respectively. The small amount of VL emitted by UV lamps (410 nm in a) was not measurable using our gauge.

Corticotropin-releasing hormone deficiency reveals major fetal but not adult glucocorticoid need

Louis Muglia*, Lauren Jacobson*, Pieter Dikkes† & Joseph A. Majzoub*

* Division of Endocrinology and

† Department of Neurology, Children's Hospital, Harvard Medical School, 300 Longwood Avenue, Boston, Massachusetts 02115, USA

THE body responds to stress by activation of the hypothalamic-pituitary-adrenal (HPA) axis and release of glucocorticoids. Glucocorticoid production in the adult regulates carbohydrate and amino-acid metabolism, maintains blood pressure, and restrains the inflammatory response¹. In the fetus, exogenous glucocorticoids accelerate maturation of lung² and gastrointestinal enzyme systems³ and promote hepatic glycogen deposition⁴. Corticotropin-

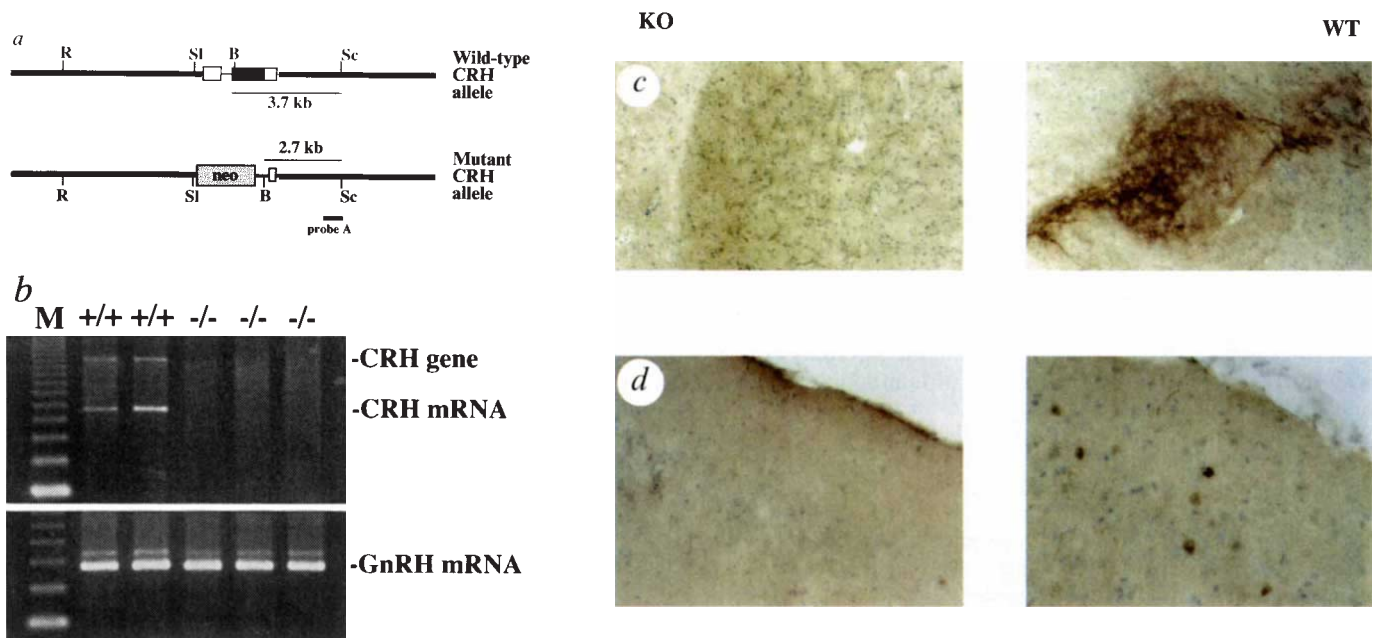


FIG. 1 CRH-deficient mice produced by targeted disruption in embryonic stem cells. **a**, Restriction maps of the normal murine CRH locus and the CRH locus resulting from targeted mutation of embryonic stem cells are shown. Homologous recombination resulted in replacement of the entire pre-proCRH coding region with the neomycin resistance gene¹⁰. The two exons of the CRH gene are indicated by the boxes on the top line, with the shaded area indicating the pre-proCRH coding region. Restriction sites indicated are *Bam*HI (B), *Eco*RI (R), *Sall* (SI) and *Sacl* (Sc). Probe A, a 350 bp *Eco*RI–*Sacl* fragment lying outside CRH genomic sequences incorporated in the targeting construct, was used to genotype mice, yielding bands of 3.7 kb for the wild-type allele and 2.7 kb for the CRH-null allele after hybridization to genomic DNA digested with *Bam*HI and *Sacl*. **b**, Reverse transcription-PCR analysis of whole brain RNA from 2 wild-type (+/+) and 3 CRH knockout (–/–) mice at embryonic day 18.5. Total RNA (2 µg) was prepared by the guanidium thiocyanate-caesium chloride method²⁷ and subjected to reverse transcription from random hexamer primers. Specific cDNAs for CRH, or gonadotropin releasing hormone (GnRH) (as a control for sample integrity), were amplified for 35 cycles using intron-spanning primers as described previously¹⁰. Amplification products were fractionated through 1.3% agarose and visualized by ethidium bromide staining of gels. The products expected for the wild-type CRH gene (1,267 bp) and mRNA (585 bp) are seen in the wild-type mice, whereas no products are evident in the knockouts. A 123 bp DNA ladder (M) was used as

releasing hormone (CRH), a 41-amino-acid neuropeptide produced in the paraventricular nucleus of the hypothalamus and many regions of the cerebral cortex^{5,6}, has been implicated in both the HPA axis⁷ and behavioural responses⁸ to stress. To define the importance of CRH in the response of the HPA axis to stress and fetal development, we have constructed a mammalian model of CRH deficiency by targeted mutation in embryonic stem (ES) cells⁹. We report here that corticotropin-releasing hormone-deficient mice reveal a fetal glucocorticoid requirement for lung maturation. Postnatally, despite marked glucocorticoid deficiency, these mice exhibit normal growth, fertility and longevity, suggesting that the major role of glucocorticoid is during fetal rather than postnatal life.

Mice heterozygous for a null CRH allele generated as shown in Fig. 1a are phenotypically normal, and fertile. Mating of heterozygous animals produces viable homozygous CRH-deficient mice at the expected frequency, with 61 (23%) –/–, 131 (49%) +/- and 74 (28%) +/+ in the first 266 mice genotyped at 4 weeks of age. CRH-deficient male and female mice, the oldest of which are now 1 year of age, are equal in size to their littermates and appear healthy without glucocorticoid replacement.

To confirm CRH deficiency, RNA and protein analyses of homozygous mutant and wild-type animals were done. Whole-

brain RNA from fetuses at 18.5 days' gestation was subjected to reverse transcription-polymerase chain reaction (RT-PCR) analysis using CRH-specific primers¹⁰. As shown in Fig. 1b, in wild-type (+/+) mice, products are detected from the CRH gene and messenger RNA, whereas in homozygous knockout animals neither product is observed. Immunohistochemical analyses of CRH expression reveals abundant CRH immunoreactivity in wild-type animals, but no staining in the homozygous deficient mice in the central nucleus of the amygdala (Fig. 1c) and cerebral cortex (Fig. 1d), sites believed important for CRH-mediated behavioural responses. Staining is also absent in the paraventricular nucleus of the hypothalamus in knockout mice (data not shown).

The effect of CRH deficiency on hypothalamic, pituitary and adrenal development was assessed in 8-week-old male mice. No significant structural differences in the hypothalamic paraventricular nucleus, the primary site of synthesis of CRH destined to stimulate the pituitary–adrenal axis, are apparent in knockout and wild-type mice (Fig. 2a). Surprisingly, arginine vasopressin (AVP) immunostaining is not increased in the CRH-deficient animals despite low corticosterone levels in basal and stimulated states (see below). Oxytocin immunostaining is also equivalent in the knockout and wild-type male mice (data not shown).

METHODS. Ten to fifteen D3²⁹ embryonic stem cells which had undergone targeted mutation of one of their CRH alleles¹⁰ were injected into C57BL/6J blastocysts at 3.5 days post-coitum⁹. Chimaeric animals resulting from injected blastocysts were mated to C57BL/6J mice, with germ-line transmission indicated by agouti coat colour in offspring. Heterozygous mice were verified by Southern blot analysis, then mated to produce homozygous CRH-deficient mice.

brain RNA from fetuses at 18.5 days' gestation was subjected to reverse transcription-polymerase chain reaction (RT-PCR) analysis using CRH-specific primers¹⁰. As shown in Fig. 1b, in wild-type (+/+) mice, products are detected from the CRH gene and messenger RNA, whereas in homozygous knockout animals neither product is observed. Immunohistochemical analyses of CRH expression reveals abundant CRH immunoreactivity in wild-type animals, but no staining in the homozygous deficient mice in the central nucleus of the amygdala (Fig. 1c) and cerebral cortex (Fig. 1d), sites believed important for CRH-mediated behavioural responses. Staining is also absent in the paraventricular nucleus of the hypothalamus in knockout mice (data not shown).

KO

WT

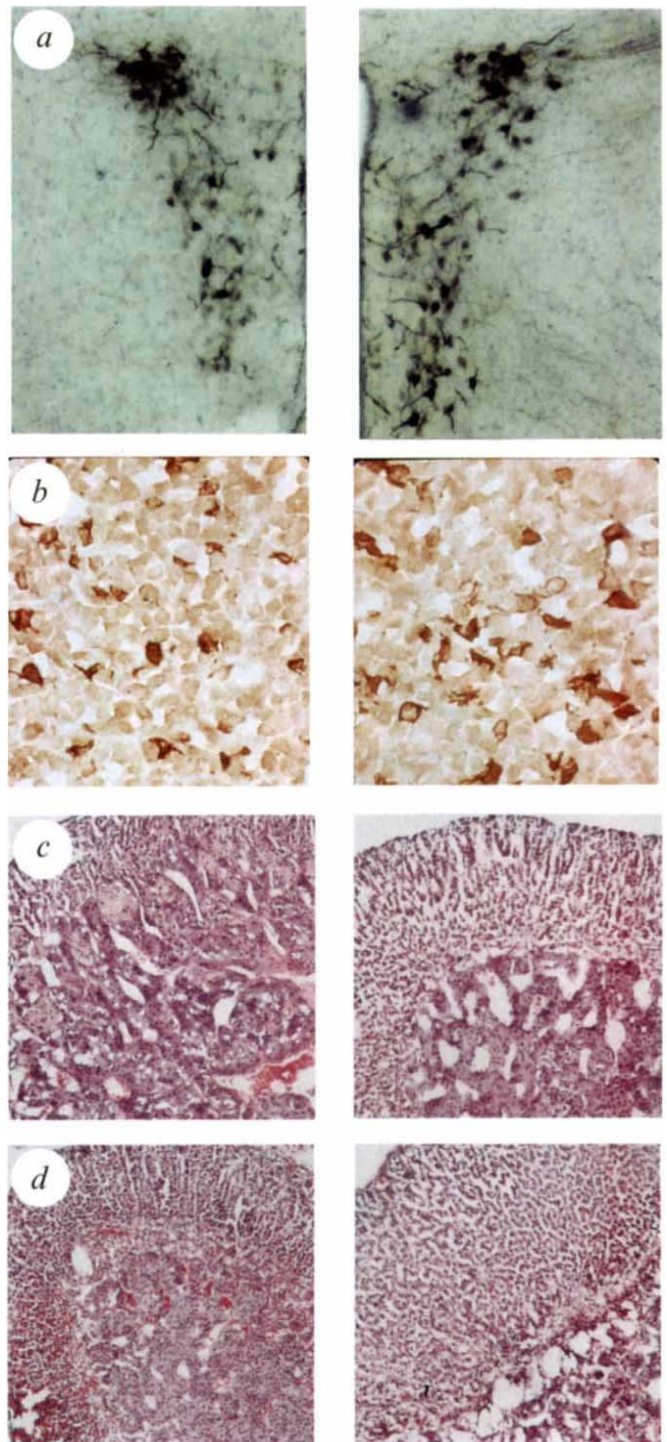


FIG. 2 Immunohistochemical analysis of the HPA axis in CRH knockout mice. *a*, Vasopressin immunostaining in the paraventricular nucleus of the hypothalamus is not increased in CRH knockout mice. Brains from 8-week-old male knockout (KO) and wild-type (WT) mice were fixed as described in Fig. 1c followed by reaction with a polyclonal vasopressin neurophysin antiserum³⁰. Antigen-antibody complexes were visualized by peroxidase staining, with anatomy defined by Nissl counterstaining. *b*, ACTH-producing corticotrophs develop in the anterior pituitary gland in the absence of CRH. Pituitaries fixed by 4% paraformaldehyde were cut into 20- μ m sections, and reacted with a rabbit polyclonal ACTH antiserum (INCSTAR Corporation, Stillwater, MN). *c*, Adrenal hypoplasia develops in CRH-deficient mice. Adrenals from 8-week-old male knockout and wild-type mice were haematoxylin-eosin stained after paraformaldehyde fixation. The characteristic strand-like appearance of the zona fasciculata is seen in the wild-type adrenal, whereas the knockout has very little zona fasciculata between the adrenal medulla (the medial, inferior region of each section) and zona glomerulosa (located beneath the adrenal capsule). *d*, CRH-deficient female mice also exhibit adrenal zona fasciculata hypoplasia, but less than that of CRH-deficient males.

CRH has been demonstrated to be mitogenic for corticotrophs *in vivo*¹¹, suggesting a role for CRH in corticotroph development. Evaluation of the anterior pituitary reveals prominent ACTH (or proopiomelanocortin) immunoreactivity both in homozygous CRH-deficient and wild-type animals (Fig. 2*b*). The number and location of positive-staining cells is similar in wild-type and knockout animals, as was immunostaining of the intermediate lobe of the pituitary. Thus, though some hypothalamic releasing factors, such as growth hormone-releasing hormone¹², function as growth or differentiation factors for their target cell type, all need not.

In contrast to the normal appearance of the pituitary is the markedly atrophic appearance of the zona fasciculata of the

adrenal gland, the area primarily responsible for corticosterone production in the mouse (Fig. 2*c, d*). Especially striking is the more atrophic appearance of the male knockout zona fasciculata in comparison with the female knockout. This is consistent with the decreased adrenal response to stress demonstrated by CRH-deficient mice in Fig. 3. The zona glomerulosa, responsible for mineralocorticoid production, appears histologically normal, consistent with the similar levels of plasma aldosterone in CRH-intact and knockout mice (25.3 ± 6.6 ng dl⁻¹ versus 24.2 ± 7.9 ng dl⁻¹, respectively, $P=0.92$). Adrenal medulla histology also appears normal.

CRH-deficient mice exhibit an impaired, sexually dimorphic adrenal response to stress (Fig. 3*a*). After restraint followed

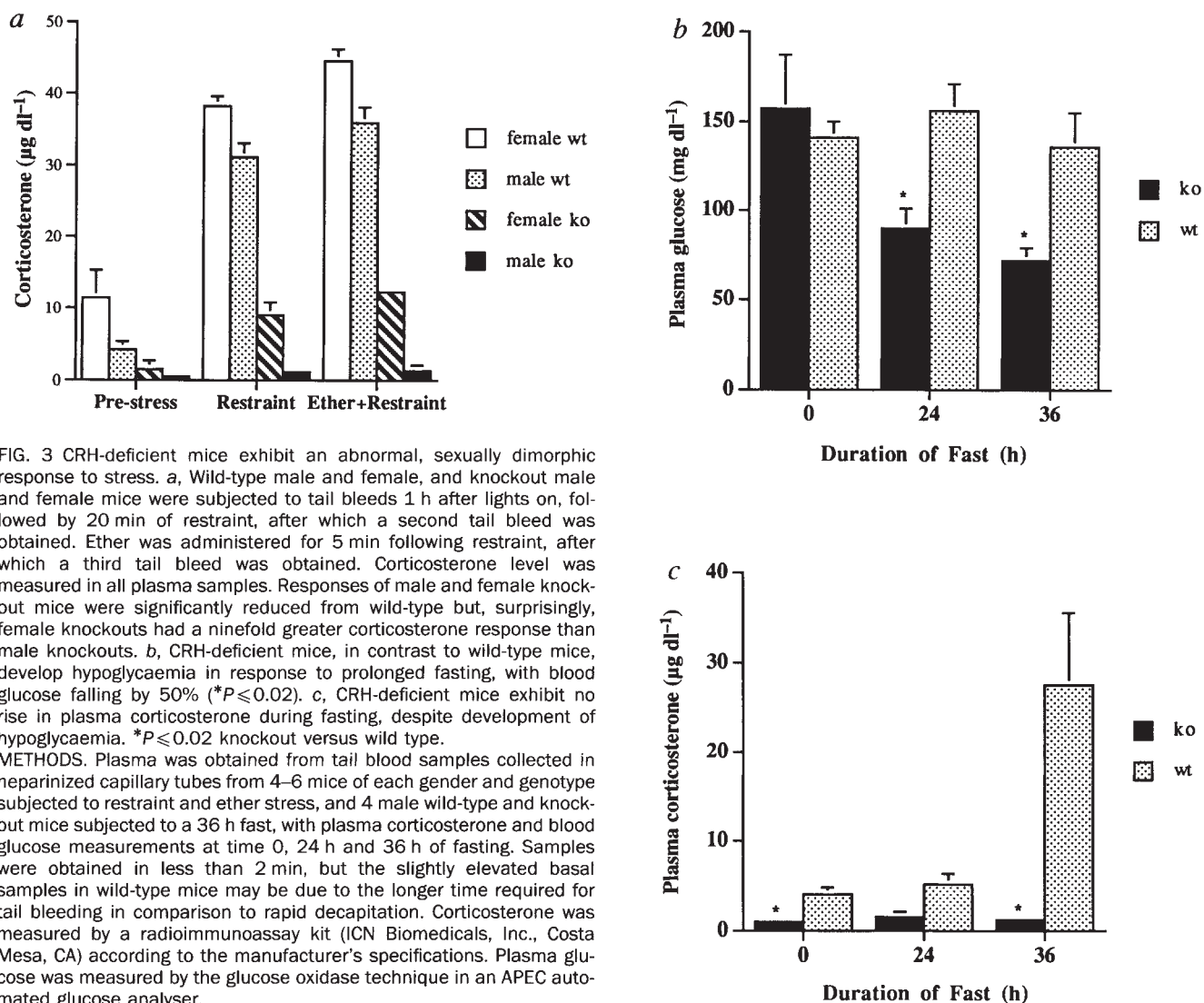


FIG. 3 CRH-deficient mice exhibit an abnormal, sexually dimorphic response to stress. **a**, Wild-type male and female, and knockout male and female mice were subjected to tail bleeds 1 h after lights on, followed by 20 min of restraint, after which a second tail bleed was obtained. Ether was administered for 5 min following restraint, after which a third tail bleed was obtained. Corticosterone level was measured in all plasma samples. Responses of male and female knockout mice were significantly reduced from wild-type but, surprisingly, female knockouts had a ninefold greater corticosterone response than male knockouts. **b**, CRH-deficient mice, in contrast to wild-type mice, develop hypoglycaemia in response to prolonged fasting, with blood glucose falling by 50% (* $P \leq 0.02$). **c**, CRH-deficient mice exhibit no rise in plasma corticosterone during fasting, despite development of hypoglycaemia. * $P \leq 0.02$ knockout versus wild type.

METHODS. Plasma was obtained from tail blood samples collected in heparinized capillary tubes from 4–6 mice of each gender and genotype subjected to restraint and ether stress, and 4 male wild-type and knockout mice subjected to a 36 h fast, with plasma corticosterone and blood glucose measurements at time 0, 24 h and 36 h of fasting. Samples were obtained in less than 2 min, but the slightly elevated basal samples in wild-type mice may be due to the longer time required for tail bleeding in comparison to rapid decapitation. Corticosterone was measured by a radioimmunoassay kit (ICN Biomedicals, Inc., Costa Mesa, CA) according to the manufacturer's specifications. Plasma glucose was measured by the glucose oxidase technique in an APEC automated glucose analyser.

by ether exposure, wild-type female mice achieve corticosterone levels of $44.5 \pm 1.7 \mu\text{g dl}^{-1}$, whereas knockout female mice achieve levels of only $12.3 \pm 0.7 \mu\text{g dl}^{-1}$ ($P < 0.0001$). In wild-type male mice, plasma corticosterone rises to $32.4 \pm 0.9 \mu\text{g dl}^{-1}$ after the combined stress, whereas in male CRH-deficient mice it is only $1.3 \pm 0.8 \mu\text{g dl}^{-1}$ ($P < 0.0001$), a value similar to the minimum level observed at the circadian nadir in normal rodents^{13,14}. Male heterozygotes exhibit a pattern of response to restraint and ether that does not differ from wild-type males (data not shown).

Because food withdrawal is a potent activator of the HPA axis¹⁵, wild-type and CRH-deficient male mice were subjected to a 36 h fast (Fig. 3b, c). During the fast, wild-type mice had no fall in plasma glucose and exhibited a significant, sevenfold rise in plasma corticosterone. In contrast, the CRH-deficient mice developed hypoglycaemia by 24 h into the fast. Despite hypoglycaemia, the knockouts had no elevation in corticosterone, which remained $< 1.5 \mu\text{g dl}^{-1}$ throughout the fast.

Despite an impaired response to stress, male and female CRH-deficient mice exhibit normal viability and fertility without glucocorticoid replacement. Matings between homozygous CRH-deficient mice result in pregnancies which deliver at 19–20 days' gestation, but yield progeny which all die within the first 12 h of life despite a normal appearance at the time of delivery. Because there is no increase in mortality of CRH-deficient animals from heterozygote matings, the heterozygous mother must provide a factor which crosses the placenta and is

capable of rescuing the homozygous offspring. CRH, present in minimal amounts during rodent (in contrast to primate¹⁶) gestation, is unlikely to cross the placenta in amounts capable of augmenting fetal adrenal activity. To investigate corticosterone deficiency during gestation as the aetiology of fetal demise, corticosterone ($30 \mu\text{g ml}^{-1}$ in the drinking water¹⁷) was administered to pregnant homozygous female mice which had been mated with homozygous males, beginning at 12 days' gestation and extending through post-partum day 14. This has uniformly resulted in the production of viable litters which have survived without subsequent postnatal glucocorticoid treatment.

Explanation for this effect of perinatal corticosterone on neonatal viability was revealed by histological examination of newborn lungs. Offspring of homozygous matings show lung dysplasia in comparison to wild-type neonates at 2 h of life (Fig. 4a, b). Marked hypercellularity of the lungs, thickened alveolar septae, and a paucity of air spaces are apparent in knockout animals compared to wild-type. Biochemical evidence of impaired lung function and increased alveolar surface tension, consistent with the abnormal histology, was obtained by analysis of surfactant apoprotein mRNA^{2,18}. Surfactant apoprotein-B mRNA, which displays glucocorticoid responsiveness in lung explants^{2,19}, is reduced to 44% of the wild-type value on embryonic day 18.5 ($P < 0.05$). Administration of corticosterone beginning at day 12 of gestation completely reverses the abnormal lung architecture in homozygous deficient mice from homo-

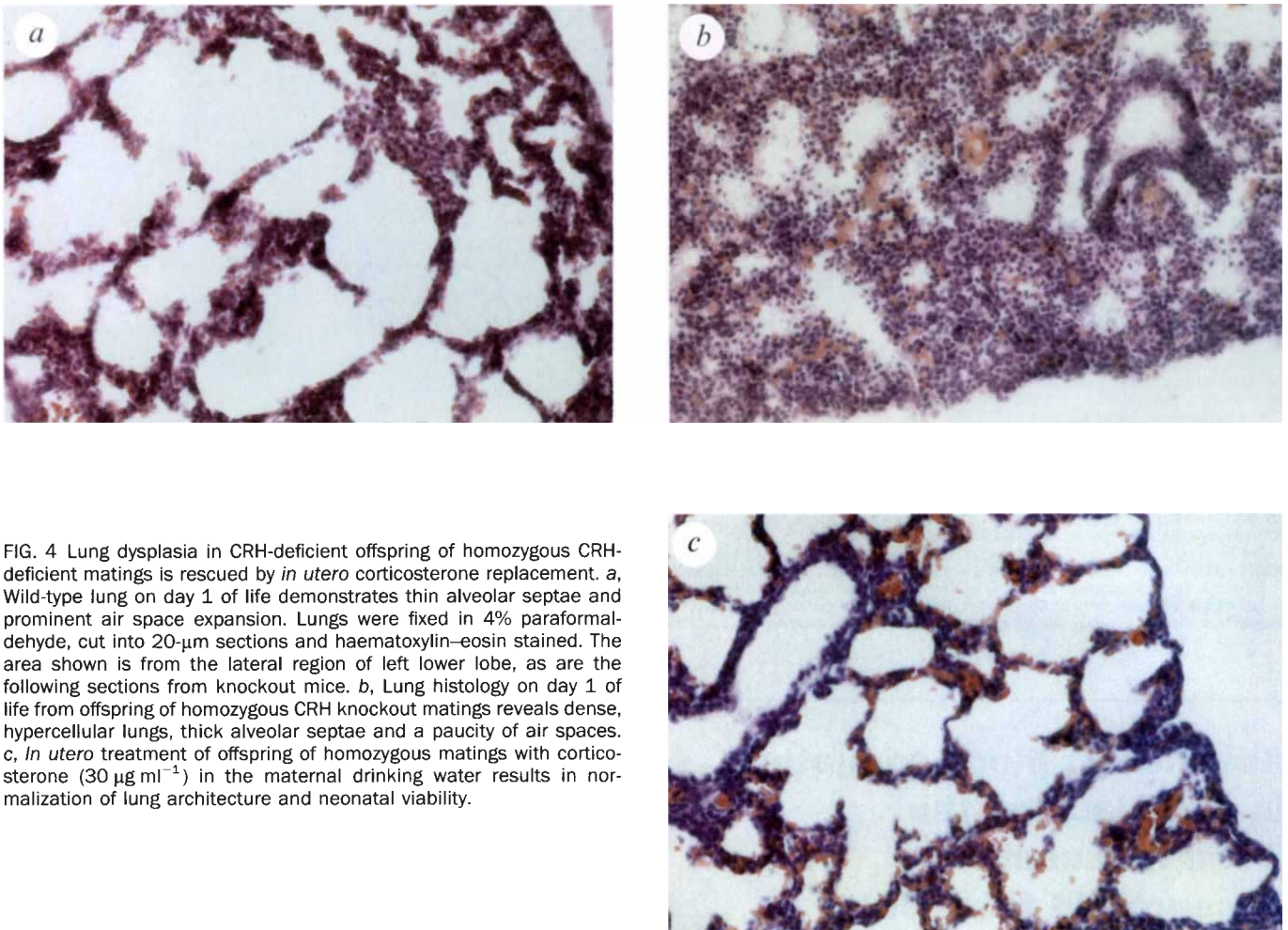


FIG. 4 Lung dysplasia in CRH-deficient offspring of homozygous CRH-deficient matings is rescued by *in utero* corticosterone replacement. *a*, Wild-type lung on day 1 of life demonstrates thin alveolar septae and prominent air space expansion. Lungs were fixed in 4% paraformaldehyde, cut into 20- μ m sections and haematoxylin-eosin stained. The area shown is from the lateral region of left lower lobe, as are the following sections from knockout mice. *b*, Lung histology on day 1 of life from offspring of homozygous CRH knockout matings reveals dense, hypercellular lungs, thick alveolar septae and a paucity of air spaces. *c*, *In utero* treatment of offspring of homozygous matings with corticosterone (30 μ g ml⁻¹) in the maternal drinking water results in normalization of lung architecture and neonatal viability.

zygous matings (Fig. 4c), and increases surfactant apoprotein-B mRNA levels at embryonic day 18.5 such that they do not differ significantly from wild-type controls (78% of wild type, $P=0.14$) but are significantly increased from untreated homozygous knockouts ($P<0.05$).

Pharmacological administration of glucocorticoids to pregnant women has long been a therapy for hyaline membrane disease, with increases in preterm infant viability after maternal administration^{20,21,22}. *In vivo* and *in vitro* studies have documented the ability of glucocorticoids to promote lung maturation^{2,19,23}. The absolute requirement for glucocorticoid as a component of normal lung development in an intact system had not been rigorously established, however. We show that both maternal and fetal circulations must be glucocorticoid deficient to cause abnormal lung development and decreased fetal viability. This may explain why children with fetal deficiencies in steroidogenesis (congenital adrenal hyperplasia) or congenital adrenal hypoplasia have normal pulmonary development despite very low glucocorticoid production.

CRH-deficient mice exhibit a significantly impaired ability to secrete corticosterone from the adrenal gland in response to stress. This demonstrates that other ACTH secretagogues cannot compensate for the loss of CRH on adrenal development or corticosterone production. One unexpected finding in these studies is the sexual dimorphism in adrenal responsiveness exhibited in CRH-deficient mice. Greater basal and stimulated corticosterone levels in normal female versus male rodents, similar in magnitude to the findings in our wild-type mice, have been demonstrated by several groups^{13,24,25}. However, the ninefold higher level of corticosterone following stress achieved by knock-

out females compared with males implicates a role for sex steroids in regulation of the HPA axis that is not mediated through CRH.

The apparent well being of the CRH-deficient mice, especially the males, contradicts the dogma of glucocorticoids being essential for adult survival. Possibly, the maximum corticosterone levels of ~ 1 –2 μ g dl⁻¹ that male CRH-deficient animals achieve (at both circadian nadir and peak, our unpublished observation) may fulfill an essential, minimal, permissive glucocorticoid requirement²⁶, especially as these mice are deficient throughout development and may have chronically accommodated in some way to the low level of glucocorticoid production. Alternatively, high CRH levels in intact animals, rather than low glucocorticoid levels, may confer many of the manifestations characteristic of adrenal insufficiency, such as anorexia, wasting, malaise, fatigue and decreased fertility. Thus, with the possible exception of stimuli more severe than those tested here, the primary importance of glucocorticoid may be in regulating fetal development, rather than supporting survival in adult life. □

Received 12 August; accepted 30 November 1994.

1. Orth, D. N., Kovacs, W. J. & Debold, C. R. in *Williams Textbook of Endocrinology* (eds Wilson, J. D. & Foster, D. W.) 489–619 (Saunders, Philadelphia, 1992).
2. Ballard, P. L. *Endocrin. Rev.* **10**, 165–181 (1989).
3. Moog, F. J. *exp. Zool.* **124**, 329–346 (1953).
4. Ballard, P. L. in *Glucocorticoid Hormone Action* (eds Baxter, J. D. & Rousseau, G. G.) 493–515 (Springer, Heidelberg, 1979).
5. Seasholtz, A. F., Bourbonnais, F. J., Harnden, C. E. & Camper, S. *Molec. cell. Neurosci.* **2**, 266–273 (1991).
6. Thompson, R. C., Seasholtz, A. F. & Herbert, E. *Molec. Endocrin.* **1**, 363–370 (1987).
7. Vale, W., Spiess, J., Rivier, C. & Rivier, J. *Science* **213**, 1394–1397 (1981).

8. Koob, G. et al. in *Corticotropin Releasing Factor* (ed. Vale, W.) 277–289 (Wiley, Chichester, England, 1993).
9. Bradley, A. in *Teratocarcinomas and Embryonic Stem Cells: A Practical Approach* (ed. Robertson, E. J.) 113–151 (IRL, Oxford, 1987).
10. Muglia, L. J., Jenkins, N. A., Gilbert, D. J., Copeland, N. G. & Majzoub, J. A. *J. clin. Invest.* **93**, 2066–2072 (1994).
11. Gertz, B. J. et al. *Endocrinology* **120**, 381–388 (1987).
12. Lin, S. C. et al. *Nature* **364**, 208–213 (1993).
13. Critchlow, V., Liebelt, R. A., Bar-Sela, M., Mountcastle, W. & Lipscomb, H. S. *Am. J. Physiol.* **205**, 807–814 (1963).
14. Dallman, M. F. et al. *Recent Prog. Horm. Res.* **43**, 113–173 (1987).
15. Dallman, M. F. et al. *Frontiers Neuroendocrin.* **14**, 303–347 (1993).
16. Sasaki, A. et al. *J. clin. Endocrin. Metab.* **59**, 812–814 (1984).
17. Jacobson, L., Akana, S., Cascio, C., Scribner, K. & Dallman, M. *Endocrinology* **124**, 2144–2152 (1989).
18. Phelps, D. S. & Floros, J. *Am. J. Physiol.* **260**, L146–L152 (1991).
19. Venkatesh, V. C., Iannuzzi, D. M., Ertsey, R. & Ballard, P. L. *Am. J. Resp. Cell Molec. Biol.* **8**, 222–228 (1993).
20. Avery, M. E. *J. Pediatr.* **104**, 240 (1984).
21. Crowley, P., Chalmers, I. & Keirse, M. J. N. C. *Br. J. Obstet. Gynecol.* **97**, 11–25 (1990).
22. Liggins, C. G. & Howie, R. N. *Pediatrics* **50**, 515–525 (1972).
23. Post, M., Barsoumian, A. & Smith, B. T. *J. Biol. Chem.* **261**, 2179–2184 (1986).
24. Vamvakopoulos, N. C. & Chrousos, G. P. *Endocrin. Rev.* **15**, 409–420 (1994).
25. Spinedi, E., Suescun, M. O., Hadid, R., Daneva, T. & Gaillard, R. *Endocrinology* **131**, 2430–2436 (1992).
26. Munck, A., Guyre, P. M. & Holbrook, N. J. *Endocrin. Rev.* **5**, 25–44 (1984).
27. Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. *Biochemistry* **18**, 5294–5299 (1979).
28. Suda, T. et al. *J. clin. Endocrin. Metab.* **59**, 861–866 (1984).
29. Doetschman, T. C., Eistetter, H., Katz, M., Schmidt, W. & Kemler, R. *J. Embryol. exp. Morphol.* **87**, 27–45 (1985).
30. Roberts, M. M. et al. *Neuroendocrinology* **57**, 388–400 (1993).

ACKNOWLEDGEMENTS. We thank R. Emanuel and R. Mortensen for assistance, F. Grant, J. Floros and S. Radovick for reagents, F. Alt for use of microinjection equipment, and members of the Majzoub laboratory for helpful discussions. This work was supported by grants from the Markey Trust, Hood Foundation, and Pfizer Pharmaceuticals to L.M., NARSAD to L.J. and the NIH to L.M. and J.M.

Absence of blood formation in mice lacking the T-cell leukaemia oncprotein tal-1/SCL

Ramesh A. Shivdasani*, Erica L. Mayer* & Stuart H. Orkin[†]*

* Division of Hematology/Oncology, Children's Hospital and the Dana Farber Cancer Institute, Department of Pediatrics, Harvard Medical School and [†] Howard Hughes Medical Institute, Boston, Massachusetts 02115, USA

CHROMOSOMAL translocations associated with malignancies often result in deregulated expression of genes encoding transcription factors¹. In human T-cell leukaemias such regulators belong to diverse protein families and may normally be expressed widely (for example, Ttg-1/rbtl, Ttg-2/rbtl^{2,3}), exclusively outside the haematopoietic system (for example, Hox11)⁴, or specifically in haematopoietic cells and other selected sites (for example, tal-1/SCL, lyl-1)^{5,6}. Aberrant expression within T cells is thought to interfere with programmes of normal maturation. The most frequently activated gene in acute T-cell leukaemias, tal-1 (also called SCL)^{7,8}, encodes a candidate regulator of haematopoietic development⁹, a basic-helix-loop-helix protein⁵, related to critical myogenic¹⁰ and neurogenic¹¹ factors. Here we show by targeted gene disruption in mice¹² that tal-1 is essential for embryonic blood formation *in vivo*. With respect to embryonic erythropoiesis, tal-1 deficiency resembles loss of the erythroid transcription factor GATA-1^{13,14} or the LIM protein rbt2¹⁵. Profound reduction in myeloid cells cultured *in vivo* from tal-1 null yolk sacs suggests a broader defect manifest at the myelo-erythroid or multipotential progenitor cell level.

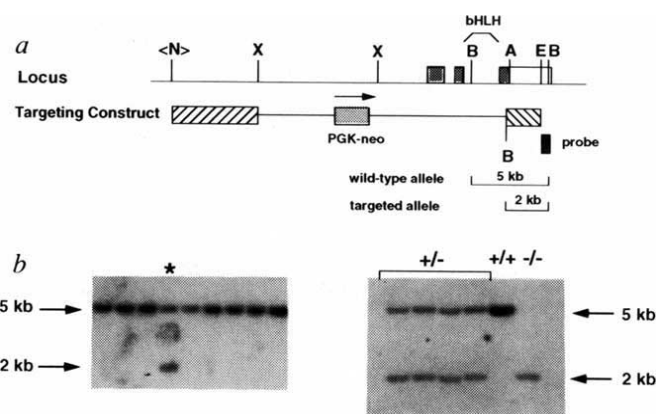


FIG. 1 Targeted disruption of the mouse tal-1 locus. a, Partial restriction map of the mouse tal-1 locus with the targeting construct depicted below. Homologous recombination resulted in the replacement of all genomic sequences lying between the two regions of homology (hatched boxes) by the PGK-neo cassette. X, XhoI; B, BamHI; A, Asp718; E, EcoRI; bHLH, basic-helix-loop-helix region; <N>, an artificial NotI site within the polylinker of the λFixII phage vector. The three terminal exons of the gene are indicated as closed boxes in top diagram. The bHLH region is encoded by the two indicated exons. The open box represents the 3'-untranslated region. A herpes virus-thymidine kinase cassette was present in the pPNT²⁴ cloning vector, outside the indicated region of homology. The wild-type (5 kb) and targeted (2 kb) BamHI fragments, as detected by a probe at the extreme 3'-end of the gene, are indicated below. b, Disruption of the tal-1 locus. Left, Southern blot analysis of ES cell clones. A targeted clone is indicated by the asterisk. Right, Southern blot analysis of E10.5 embryos resulting from interbreeding of tal-1^{+/-} mice.

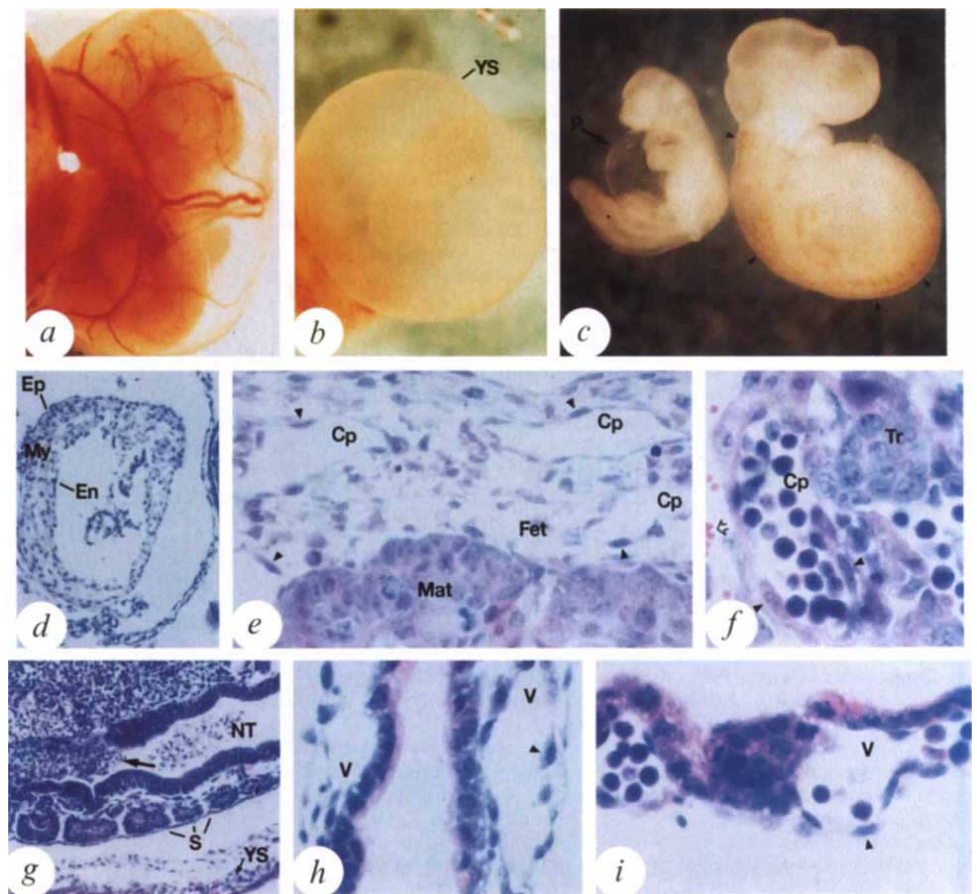
METHODS. Overlapping genomic phage clones of the tal-1 locus were isolated from a λFixII strain 129 library (Stratagene) by standard procedures²⁵. Partial restriction mapping, PCR analysis, and limited DNA sequencing were done to identify exons, particularly those encoding the bHLH domain²⁶. The 5-kb 5'-(<N>-X) and 2-kb 3'-(A-E) fragments were cloned into the vector pPNT²⁴ to generate the targeting construct. The 3'-flanking probe was generated by PCR labelling using a specific primer pair (5'-TGTTCTCATCTCCATACCCC-3'/5'-AAGTACGGCTAGACCC-ACCAA-3') and a 3'-untranslated region fragment as template. The targeting construct was linearized with NotI and electroporated into J1 ES cells as described^{27,28}. Clones were selected in G418 (250 µg ml⁻¹) and ganciclovir (2.5 µM), expanded and subjected to Southern blot analysis using the 3'-probe. Homologous recombination was observed at a frequency of ~1/200 G418/ganciclovir-resistant clones. Chimaeras were generated by injection of C57BL/6 blastocysts as previously described²⁸. Genotyping of liveborn offspring and embryos was done by either Southern blot or PCR analysis.

To disrupt the tal-1 gene, we constructed a targeting vector in which a neomycin-resistance cassette replaced coding sequences (Fig. 1a). Homologous recombination in transfected embryonic stem (ES) cells was detected by Southern blot analysis (Fig. 1b). Heterozygous mice (tal-1^{+/-}) appeared normal and were bred together. Among 76 pups from 12 litters no liveborn tal-1^{-/-} mice were observed. At embryonic day 8.5 (E8.5) ~25% of viable embryos were of the tal-1^{-/-} genotype; none was identified after E10.5. Thus, loss of tal-1 is lethal between E8.5 and E10.5. The tal-1^{-/-} embryos lacked genomic sequences represented in tal-1 messenger RNA (data not shown).

E9.5–10 tal-1^{-/-} embryos were readily distinguished from littermates by their smaller size, profound pallor, and a dilated pericardial sac (Fig. 2a–c). Although overall growth of tal-1^{-/-} embryos was retarded, major developmental milestones, including chorio-allantoic fusion, rotation of the embryo, formation of head structures, closure of the caudal neural pore, and initiation of cardiac contractions, were attained. Further visual inspection revealed that tal-1^{-/-} embryos were bloodless. Histological analysis established the absence of embryonic (nucleated) red cells in the embryo, yolk sac and placenta but demonstrated

[†] To whom correspondence should be addressed.

FIG. 2 Phenotype of *tal-1*^{-/-} embryos. Wild-type (a) and *tal-1*^{-/-} (b) E10.0 embryos with intact visceral yolk sacs (YS) are shown. Blood-filled vessels in the yolk sac and embryo are seen in the littermate control (a) but not in the mutant (b). When dissected free of the yolk sac, *tal-1*^{-/-} embryos (c, left) at E9.5 are further distinguished from normal littermates (c, right) by smaller overall size, dilated pericardial sac (P), and absence of intravascular blood (arrowheads). At this stage, *tal-1*^{-/-} embryos have an average of 21 somites compared with ≥ 25 in normal littermates. d, Transverse section through the heart of a *tal-1*^{-/-} E9.5 embryo, demonstrating the presence of epicardial (Ep), myocardial (My) and endocardial (En) layers and absence of intraluminal nucleated blood cells. In many *tal-1*^{-/-} embryos, as shown, cells with an appearance consistent with endocardial origin appeared to have collapsed within the cardiac lumen. e and f, Transverse sections through the placentas of *tal-1*^{-/-} (e) and control (f) E9.5 embryos, demonstrating absence of embryonic blood cells within endothelial-lined (arrowheads) capillaries in the fetal labyrinth (Fet) of *tal-1*^{-/-} placentas. Maternal, non-nucleated mature red cells were identified within sinusoids on the trophoblastic (Mat) side in both cases (open arrow in f). Although chorio-allantoic fusion was evident in the *tal-1*^{-/-} placenta (e), interdigitations between trophoblastic (Tr) and allantois-derived elements had not formed as in controls (f). g, Parasagittal section through mid-portion of a *tal-1*^{-/-} E9.5 embryo, demonstrating formation of somites (S) and neural tube (NT), and extensive necrosis (arrow) within the adjacent primitive mesenchyme and portions of the neural tube. Some *tal-1*^{-/-} embryos had a grossly visible plug corresponding to the necrosis in this region, caudal to the otic placodes. h, i, Sections through the visceral yolk sacs of *tal-1*^{-/-} (h) and control (i) E9.5 embryos, demonstrating similar appearance of



yolk sac tissue and endothelial-lined (arrowheads) blood vessels (V) in each case, but absence of embryonic blood cells in the mutant. (Original magnifications: d, g, $\times 100$; e, f, h, i, $\times 400$; Figure reduction, 66%.)

METHODS. Embryos of the stated age were dissected free of maternal tissues and photographed, then fixed in 10% buffered formalin and embedded in paraffin. Sections (5 μ m) were cut and stained with haematoxylin and eosin for histological examination.

the presence of other tissues of mesodermal origin, including somites, blood vessels and myocardium (Fig. 2d-i). The evident growth retardation, extensive necrosis (Fig. 2c, g), and subsequent death probably result from extreme anaemia coinciding with rapid fetal growth.

To determine molecular correlates of *tal-1* deficiency we did reverse transcriptase-polymerase chain reaction (RT-PCR) analysis of RNA obtained from yolk sac tissue (Fig. 3). In accord with the nature of the targeted mutation, no *tal-1* transcripts were detected. GATA-1¹⁶, *c-myb*¹⁷ and embryonic globin RNAs were also undetectable. In contrast, transcripts for actin (not shown), α -fetoprotein (a marker for yolk-sac endoderm) and *rbt2* (a widely expressed LIM protein required for erythroid development¹⁵), were present at the same levels as in littermate controls. Taken together with the histological findings, these data indicate that the observed phenotype results from a complete and early block in embryonic erythropoiesis, rather than from maturational arrest of a precursor population.

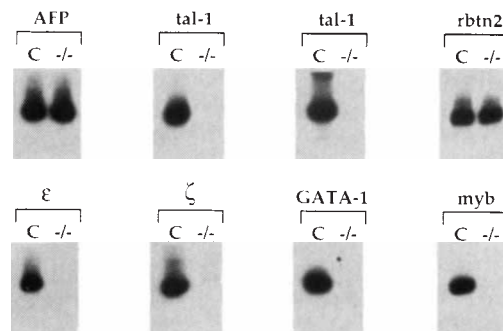
Although the absence of nucleated red blood cells in *tal-1*^{-/-} embryos reflects a failure to produce primitive erythroid cells, it does not reveal the effects on definitive (fetal liver stage) haematopoiesis or the extent of other haematopoietic lineages affected. To address these aspects, we assayed for haematopoietic progenitor cell colonies with cells from E8.0–9.5 yolk sacs. Under assay

conditions in which the majority of colonies obtained from normal yolk sac cells are of a definitive type¹⁸, *tal-1*^{+/-} and *tal-1*^{+/+} yolk sac cells generated numerous erythroid, myeloid (macrophage/granulocyte) and mixed colonies (Fig. 4). In contrast, *tal-1*^{-/-} yolk sac cells yielded scattered yolk sac vesicles and numerous fibroblasts, but no primitive, definitive or developmentally arrested erythroid colonies. Exceedingly rare (<1 per 5 *tal-1*^{-/-} yolk sacs), small clusters of macrophage-like cells were observed only after prolonged culture (15–20 days). This virtual absence of myeloid colony formation is more likely to reflect a deficit at the myelo-erythroid progenitor level than a trivial effect of extreme anaemia, because E9.5 *rbt2*-null mice, which appear comparably anaemic, were not reported to be markedly deficient in macrophage progenitors *in vitro*¹⁵.

Our findings provide strong evidence for a fundamental role for *tal-1* in embryonic haematopoiesis, analogous to the inferred function of related basic-helix-loop-helix (bHLH) factors in myogenesis and neurogenesis^{10,11,19}. Surprisingly, this central role of *tal-1* protein is not compensated for by *lyl-1* protein⁶, another leukaemia-associated factor which shares ~80% identity in the bHLH region and an overlapping pattern of expression among haematopoietic cell lines⁹. The extreme effect of *tal-1* loss on early aspects of erythropoiesis does not exclude a potential role for the protein during later erythroid maturation, as previous studies in cultured cell lines have implicated *tal-1* as a

FIG. 3 RT-PCR analysis of yolk sac RNA from *tal-1*^{-/-} and control (C, *tal-1*^{+/+}) E9.5 embryos. The presence of transcripts corresponding to α -fetoprotein (AFP), *tal-1* protein (using two separate primer pairs), *rbtn2*, embryonic ϵ - and ζ -globins, and transcription factors GATA-1 and *c-myb*, was tested in *tal-1*^{-/-} and control samples.

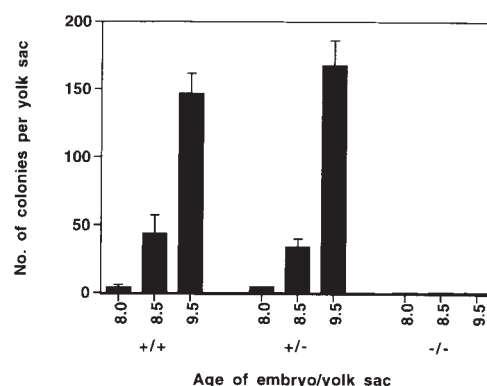
METHODS. Yolk sacs from E9.5 embryos were carefully dissected free of maternal tissues using fine forceps, and RNA prepared by the one-step method²⁹. First-strand cDNA was synthesized using M-MLV reverse transcriptase (Boehringer-Mannheim) at 42 °C for 1 h, and 1 μ l of this reaction mixture was used directly in 25- μ l PCR reactions (94 °C $\times$ 1 min, 60 °C $\times$ 1 min, 72 °C $\times$ 2 min) with tracer ³²P-dCTP, as in ref. 14. All PCR reactions were done for a number of cycles corresponding to the high end of the range in which a linear increase in products could be detected: 24 cycles for globins and 36 cycles for others. No reaction products were detected using RNA samples to which no reverse transcriptase was added (not shown). Reaction products were analysed on non-denaturing 4% polyacrylamide gels, which were dried and subjected to autoradiography. Primer sets: ϵ - and ζ -globins and *c-myb*, as in ref. 14; α -fetoprotein, as in ref. 30; *tal-1*, 5'-GTCTCAC-ACCAAGTAGTG-3' or 5'-AACTAAGCAAGAATGAGATC-3' (forward) and



5'-GGCACCTCAAAGCTTGACTCTCCA-3' (reverse) yielding products of 984 bp and 861 bp, respectively; *rbtn2*, 5'-TCGCGAATTCACCATGTCCTCGGCCATCGAAAGGA-3' and 5'-CGAGTCTAGACTAGATGATCCCATGATCTTGGT-3' (product ~450 bp); and GATA-1, 5'-CATTGGCCCCCTTG-TGAGGCCAGAGA-3' and 5'-CGGAGATAAAGTTCGAGGTAGTCCA-3' (product ~280 bp).

FIG. 4 Haematopoietic progenitor colony assays from yolk sacs of *tal-1*^{-/-} embryos and littermates. The total number of haematopoietic colonies observed after 8 days of culture of entire yolk sacs is shown from E8.0–E9.5 embryos with *tal-1*^{+/+}, *tal-1*^{+/-}, and *tal-1*^{-/-} genotypes. The majority of colonies from *tal-1*^{+/+} and *tal-1*^{+/-} yolk sacs contained cells corresponding to a mixture of morphological cell types, including non-erythroid cells; between 60 and 80% of the colonies included visibly haemoglobinized cells. Results from two separate experiments are shown, and expressed as the mean $\pm$ s.e.m.

METHODS. Yolk sacs from embryos of the indicated age were carefully dissected free of maternal tissues under sterile conditions. E8.0 and E8.5 yolk sacs were disaggregated by passage through a syringe and 22-gauge needle; E9.5 yolk sacs were incubated in Ca²⁺/Mg²⁺-free phosphate-buffered saline (PBS) with 20% fetal bovine serum (FBS) and 0.1% collagenase (Sigma) at 37 °C for 30 min before mechanical disaggregation¹⁸. The yield from each *tal-1*^{+/+} and *tal-1*^{+/-} yolk sac was 4×10^4 to 1×10^5 cells; $<10^3$ such cells, presumably largely of non-haematopoietic origin, were obtained from *tal-1*^{-/-} yolk sacs. After washing once in PBS, cells were plated in α -minimal essential medium supplemented with 0.9% methylcellulose, 30% FBS, 1% BSA, 2 units ml⁻¹ erythropoietin, 50 ng ml⁻¹ recombinant c-kit ligand (Amgen), and



0.5% pokeweed mitogen-stimulated murine spleen cell conditioned medium (Stem Cell Technologies). Replacement of erythropoietin by recombinant murine GM-CSF (20 U ml⁻¹ did not increase *tal-1*^{-/-} colony yield (not shown).

positive regulator of erythroid maturation²⁰. Although *tal-1* has been detected in progenitors of both haematopoietic and vascular tissues in the early mouse embryo and adult^{21,22}, our data reveal its requirement in blood cell development but not in vasculogenesis. Embryonic lethality precludes simple study of the role of *tal-1* at additional sites of limited expression, such as the central nervous system and thymus.

Taken together with earlier independent findings, the *tal-1*^{-/-} phenotype strengthens the interrelationship between *tal-1*, GATA-1 and *rbtn2*. First, one of two promoters of the *tal-1* gene is activated by a GATA-factor²⁰; hence, at some developmental stages *tal-1* may be a downstream target of GATA transcription factors. Second, the *tal-1* and GATA-1 genes share a remarkably similar pattern of expression⁹, and high levels of *rbtn2* are found in erythroid cells¹⁵. Third, *tal-1* protein and *rbtn2* physically associate within the nuclei of erythroleukaemia cells²³, which suggests that *rbtn2* might influence the DNA binding and/or transactivation properties of *tal-1*. Finally, the consequence for embryonic erythroid development is very similar on disruption of each of the three genes^{13,15}. However, in contrast to GATA-1 and *rbtn2* deficiencies, the scarcity of myeloid colonies generated *in vitro* from *tal-1*^{-/-} yolk sac cells (Fig. 4) suggests that *tal-1* may be required within a myelo-erythroid or multipotential progenitor. Thus, the precise hierarchy of GATA-1, *tal-1*, and *rbtn2* within haematopoiesis remains to be more fully defined. On the basis of the findings reported here, we speculate that genes activated by *tal-1* during leukaemogenesis include those normally regulated by this factor in early haematopoietic cells. □

Received 8 November; accepted 15 December 1994.

1. Rabbitts, T. H. *Nature* **372**, 143–149 (1994).
2. Boehm, T. et al. *EMBO J.* **7**, 385–394 (1988).
3. Royer-Pokora, B., Loos, U. & Ludwig, W.-D. *Oncogene* **6**, 1887–1893 (1991).
4. Roberts, C. W. M., Shutter, J. R. & Korsmeyer, S. J. *Nature* **368**, 747–749 (1994).
5. Begley, C. G. et al. *Proc. natn. Acad. Sci. U.S.A.* **86**, 10128–10132 (1989).
6. Mellentin, J. D., Smith, S. D. & Cleary, M. L. *Cell* **58**, 77–83 (1989).
7. Brown, L. et al. *EMBO J.* **9**, 3343–3551 (1990).
8. Aplana, P. D. et al. *Science* **250**, 1426–1429 (1990).
9. Visvader, J., Begley, C. G. & Adams, J. M. *Oncogene* **6**, 187–194 (1991).
10. Olson, E. N. & Klein, W. H. *Genes Dev.* **8**, 1–8 (1994).
11. Guillemot, F. et al. *Cell* **75**, 463–476 (1993).
12. Capecchi, M. R. *Science* **244**, 1288–1292 (1989).
13. Pevny, L. et al. *Nature* **349**, 257–260 (1991).
14. Weiss, M. J., Keller, G. & Orkin, S. H. *Genes Dev.* **8**, 1184–1197 (1994).
15. Warren, A. J. et al. *Cell* **78**, 45–57 (1994).
16. Tsai, S. F. et al. *Nature* **339**, 446–51 (1989).
17. Mucenski, M. L. et al. *Cell* **65**, 677–689 (1991).
18. Wong, P. M. C., Chung, S. W., Chui, D. H. K. & Eaves, C. J. *Proc. natn. Acad. Sci. U.S.A.* **83**, 3851–3854 (1986).
19. Rudnicki, M. A. et al. *Cell* **75**, 1351–1359 (1993).
20. Aplana, P. D., Nakahara, K., Orkin, S. H. & Korsmeyer, S. J. *EMBO J.* **11**, 4073–4081 (1992).
21. Kallianpur, A. R., Jordan, J. E. & Brandt, S. J. *Blood* **83**, 1200–1208 (1994).
22. Hwang, L.-Y., Siegelman, M., Davis, L., Oppenheimer-Marks, N. & Baer, R. *Oncogene* **8**, 3043–3046 (1993).
23. Valge-Archer, V. E. et al. *Proc. natn. Acad. Sci. U.S.A.* **91**, 8617–8621 (1994).
24. Tybulewicz, V. L. J., Crawford, C. E., Jackson, P. K., Bronson, R. T. & Mulligan, R. C. *Cell* **65**, 1153–1163 (1991).
25. Ausubel, F. M. et al. *Curr. Prot. molec. Biol.* (1987).
26. Begley, C. G. et al. *Proc. natn. Acad. Sci. U.S.A.* **88**, 869–873 (1991).
27. Li, E., Bestor, T. H. & Jaenisch, R. *Cell* **69**, 915–926 (1992).
28. Tsai, F.-Y. et al. *Nature* **371**, 221–226 (1994).
29. Chomczynski, P. & Sacchi, N. *Analyt. Biochem.* **162**, 156–159 (1987).
30. Keller, G., Kennedy, G., Papayannopoulou, T. & Wiles, M. V. *Molec. cell. Biol.* **13**, 472–486 (1993).

ACKNOWLEDGEMENTS. We thank S. Goff, K. Rockwell, C. Browne and M. Rosenblatt for technical assistance, H. Warren, E. Meluleni and F. Kuo for assistance with histological analysis and K. Vignali for extraordinary patience and efforts during early phases of the project. This work was supported in part by funds from the NIDDK-NIH to establish a Center for Excellence in Hematology at the Children's Hospital. R.A.S. was supported in part by a fellowship from the American Cancer Society. S.H.O. is an Investigator of the Howard Hughes Medical Institute.

Quantal-like current fluctuations induced by odorants in olfactory receptor cells

Anna Menini, Cristiana Picco & Stuart Firestein*

Istituto di Cibernetica e Biofisica, Consiglio Nazionale delle Ricerche, 16146 Genova, Italy

* Department of Biological Sciences, Columbia University, New York, New York 10027, USA

MANY sensory systems have evolved signal detection capabilities that are limited only by the physical attributes of the stimulus¹. For example, 'hair' cells of the inner ear can detect displacements of atomic dimensions². Likewise, both in vertebrates and in invertebrates photoreceptors can detect a single photon^{3,4}. The olfactory stimulus also has a quantal unit, the single odorant molecule. Insects are reportedly able to detect a single pheromone molecule⁵, whereas quantal responses in vertebrate olfactory receptor cells have not been reported yet. Psychophysical measurements indicate that a minimum of 50 odorant molecules are necessary for human olfactory detection, suggesting that an individual receptor may be activated by a single odorant molecule⁶. We report here measurements of current fluctuations induced by odorants that suggest a quantal event of about 0.3–1 pA, presumably triggered by the binding of a single odorant molecule.

Figure 1a illustrates current recordings from an isolated receptor neuron elicited by long exposures (>30 s) to the odorant cineole at concentrations ranging from 5×10^{-7} to 7×10^{-5} M. High odorant concentrations produced a large initial current transient (up to 920 pA at 7×10^{-5} M) which decreased within a few seconds to a much smaller plateau level characterized by very pronounced fluctuations that disappeared on removal of the odorant. In the absence of odorants the current trace is quiet, as shown in the upper trace of Fig. 1b. At the lowest odorant concentrations, 5×10^{-7} and 1×10^{-6} M, no initial transient peak was observed and the response to odorant was composed of fluctuations arising from the baseline level were observed. They had amplitudes of about 1 pA but occasionally even a large fluctuation

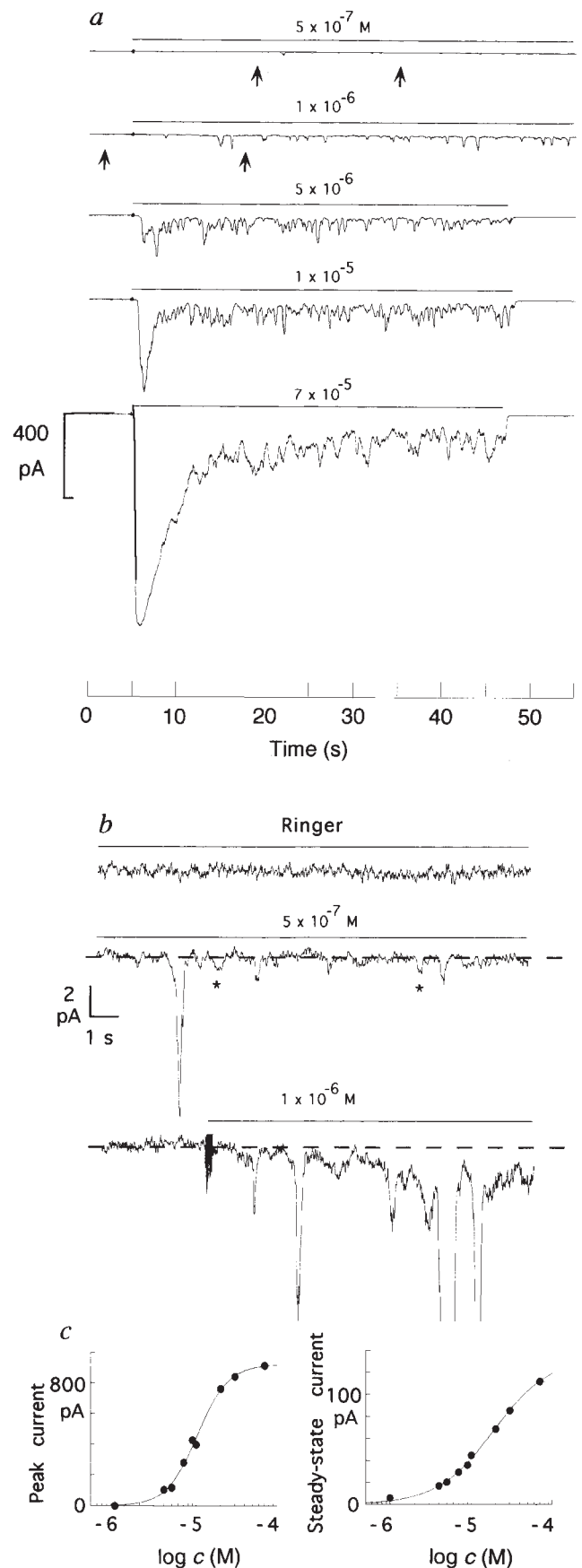


FIG. 1 Responses of an isolated olfactory receptor cell to increasing concentrations of the odorant cineole for long exposure times. a, Lines and numbers above traces indicate, respectively, duration and concentration of the stimulus. Holding potential $V_h = -50$ mV, filter 100 Hz. b, Recording obtained in Ringer's solution, in the absence of odorants and expanded sections from traces shown in a (between the arrows) digitally filtered at 20 Hz. Some fluctuations of about 1 pA are marked by an asterisk. Dashed lines indicates baseline level. c, Dose-response relations plotted as the peak current or as the steady-state average current versus odorant concentration from the cell shown in a. Continuous lines are the best fit of the Hill equation ($I = I_{max} c^n / (c^n + K_{1/2}^n)$) to the data with the following values: $I_{max} = 921$ pA, $K_{1/2} = 1 \times 10^{-5}$ M, $n = 2.3$ for the peak current; $I_{max} = 135$ pA, $K_{1/2} = 1.9 \times 10^{-5}$ M, $n = 1.2$ for the steady-state current.

METHODS. Methods for obtaining isolated olfactory receptor cells, rapid solution changes and current recordings were previously described⁷. Briefly, olfactory receptor cells were mechanically isolated from the olfactory epithelia of the terrestrial salamander (*Ambystoma tigrinum*) and currents were recorded in the whole-cell voltage-clamp mode. The intracellular pipette solution contained (in mM): CsCl, 120; MgCl₂, 2; CaCl₂, 1; EGTA, 4; HEPES, 4; ATP, 1; GTP, 0.1; pH 7.5. Odorant solutions (cineole or cinnamaldehyde) were prepared in amphibian Ringer's solution with less than 0.1% DMSO. A rapid perfusion device was used for precise control of both the exposure time and the concentration of

odorant at the cell surface. A single cell could be exposed to any of 48 different solutions. The solution switching time was 30–50 ms, as measured by the response of the cell to 20 mM KCl. All experiments were done at room temperature, $\sim 23^\circ\text{C}$.

of 10 pA was measured. Increasing the odorant concentration to 1×10^{-6} M caused an increase in the size and frequency of fluctuations. At this higher concentration, current fluctuations were rarely isolated and often ran together, emerging from a current background of a few pA. Therefore the large fluctuations shown in Fig. 1a and b do not appear to be isolated events. Returning to Ringer's solution reestablished the original baseline level. Dose-response curves measured at both the peak and the steady-state portions of the responses from the experiment in Fig. 1a are shown in Fig. 1c. The data were well fitted by the Hill equation with $K_{1/2} = 1 \times 10^{-5}$ M and $n = 2.3$ for the peak currents, and $K_{1/2} = 1.9 \times 10^{-5}$ M and $n = 1.2$ for the average steady-state currents.

In the cell shown in Fig. 1 there was no macroscopic current response below 3×10^{-6} M, although current fluctuations were observed at 5×10^{-7} M, that is almost two orders of magnitude lower than the $K_{1/2}$ for this cell. These results indicate that an isolated olfactory receptor cell in the presence of an odorant concentration much lower than its $K_{1/2}$ could produce current fluctuations of a few pA, provided that the cell is exposed to 'low odorant concentrations' for a long time. As we have previously shown⁷, olfactory receptor cells have $K_{1/2}$ values, measured at the peak of the response, ranging from 3×10^{-6} to 9×10^{-5} M and a narrow dynamic range. As a consequence there is no unique definition of low concentration that will apply to all cells and the notion of low concentration becomes relative to the $K_{1/2}$. In one cell exposed for 6 s to a concentration of cinnamaldehyde as low as 10^{-10} M we could measure current fluctuations ranging from 1 to 5 pA.

In two cells the chloride channel blocker SITS at 2 mM was added to determine the contribution of the recently characterized calcium-activated chloride current^{8,9}. In both cases the transient macroscopic current was reduced in amplitude but the fluctuations were unaffected. It appears that the fluctuations are due entirely to the cAMP-dependent current.

Experiments in which another cell was exposed to 10 consecutive steps of odorants of 6 s duration at two different concentrations are shown in Fig. 2a (3×10^{-5} M cineole) and b (1×10^{-5} M cineole). Addition of the individual traces obtained with 3×10^{-5} M cineole (bottom trace of Fig. 2a) produced a macroscopic current with a clear peak followed by a noisy plateau,

similar to the trace at 1×10^{-5} M cineole in the cell of Fig. 1a. At the lower concentration, 1×10^{-5} M cineole (Fig. 2b), fluctuations occurred at random times after the onset of the odorant step and the summed current did not show a typical macroscopic odorant-induced current, but was more similar to traces shown in Fig. 1a at 1×10^{-6} M or 5×10^{-6} M cineole. Small fluctuations of about 1 pA amplitude were also observed in these recordings. These observations suggest that the odorant responses in Fig. 1a may be composed of the superposition of many smaller events.

The small current fluctuations observed in the presence of a steady low odorant concentration (middle trace in Fig. 1b) are reminiscent of quantal fluctuations in response to dim steady illumination in rod photoreceptors³ suggesting that they also may result from the quantal nature of the odorant stimulus. To find the size of the elementary events, olfactory receptor cells were exposed to brief odorant pulses. Figure 3a shows the responses of a cell to consecutive 500-ms steps of 1×10^{-5} M cineole delivered at 6-s interstimulus intervals. Current responses showed considerable variation in amplitude. In some of the trials there was no apparent response, in others a response as small as 0.3 pA, and occasionally large responses of 2, 4 and even one response of 9.5 pA. The response amplitudes were estimated after the method of ref. 3 by scaling the amplitude of the average response to obtain the best least-squares fit to each response (Fig. 3b). In spite of the different sizes of event amplitudes all the responses were well fitted by the scaled average response, indicating that all these responses have the same characteristic kinetics. The quantal-like responses of various amplitudes and the occasional failures suggest that on average very few receptors have been activated by the odorant molecules. The histogram of event amplitudes is shown in Fig. 3c. The inset shows the histogram of interleaved trials in Ringer's solution to estimate the dispersion of the baseline. Visual inspection of the total histogram shows a component around zero (corresponding to failures) with dispersion consistent with the histogram obtained in the absence of odorants, a component with amplitudes between 0.3 and 1 pA and some isolated larger responses varying between 1 and 10 pA. Histograms from two other cells gave similar results. The histogram of Fig. 3c suggests that the size of the elementary event is between 0.3 and 1 pA and that the size of multiple events is rather variable.

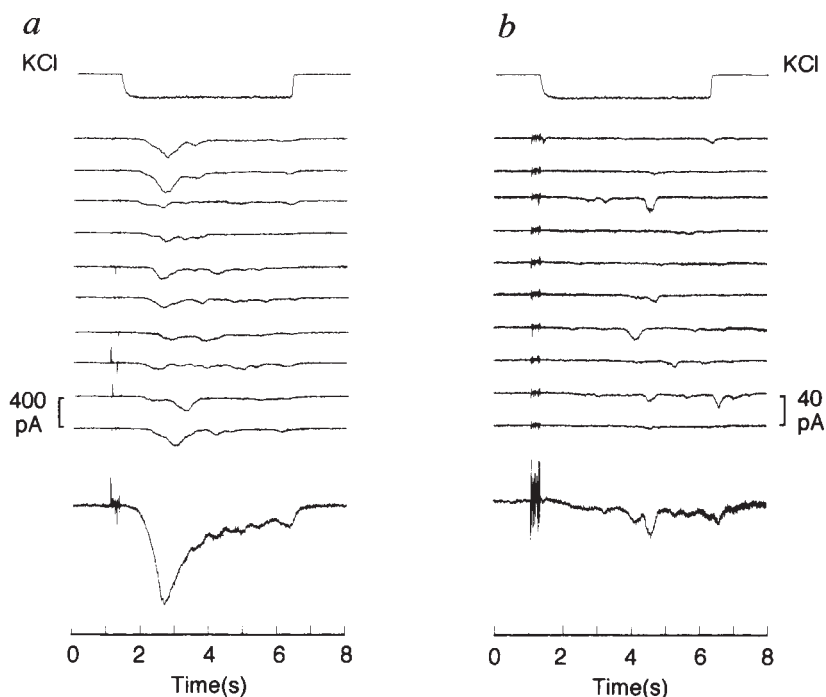


FIG. 2 Responses to repeated 6 s steps of cineole at 3×10^{-5} M (a) or 1×10^{-5} M (b). Each trial was separated by 90 s. The top traces, labelled KCl, represent the time course of the stimulus delivered to the cell, as monitored by the response of the cell to 20 mM KCl instead of odorants. An electrical artefact, caused by the stepping motor that brought the cell into the stimulus stream was present in many traces. The bottom traces are the sum of the individual responses. $V_h = -50$ mV, filtered at 100 Hz.

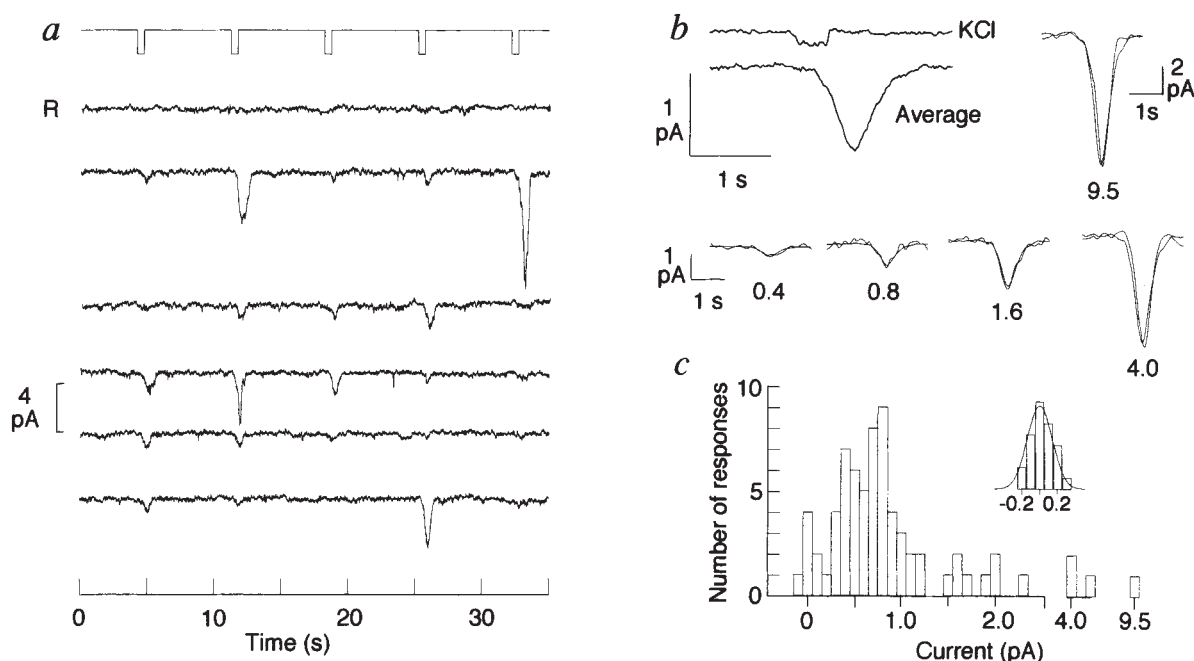


FIG. 3 Responses to repeated exposures to short pulses of odorant illustrating the quantal nature of the odorant response. *a*, Continuous recording from an isolated cell exposed to 500 ms steps of 1×10^{-5} M cineole delivered as indicated by the downward deflections of the top trace. The trace labelled R was obtained by moving the cell from Ringer into Ringer's solution and shows that there was no detectable current artefact due to the movement. $V_h = -50$ mV, filter at 50 Hz. *b*, The average of 70 responses from the cell in *a* had a peak at 1 pA. The trace labelled KCl represents the stimulus time course. The average response was then used to compute the amplitude of each response as in ref. 3.

The variability of the size of multiple events might originate from the same mechanism that we suggest is responsible for the steep relation between odorant concentration and amplitude of the macroscopic response. The morphology of the olfactory neuron with its many fine ($<1 \mu\text{m}$ diameter) individual cilia could result either in the simultaneous occurrence of several spatially independent or interacting transduction events. When odorant molecules interact with spatially isolated odorant receptors, for example on different cilia, the total current will be the linear superposition of the individual current events. However, when odorant molecules interact with spatially close odorant receptors in the same cilium, the total current is not expected to be the linear superposition of the individual events because it is activated by the internal transmitter cAMP in a cooperative way with a Hill coefficient of about $2^{10,11}$.

In vertebrate photoreceptors, where transduction is regulated

The average was scaled to provide a least-squares fit to each individual response. Some responses, digitally filtered at 3 Hz, are shown with the least-squares fit to the average. Numbers below traces are the scaling factors obtained from the fit. The scaled average provided a good fit even though individual responses had many different amplitudes, ranging from 0.3 to 9.5 pA. *c*, Amplitude histogram of 70 responses computed with the least-squares fit method from the cell shown in *a*. Inset shows histogram for recordings in the absence of odorants obtained by scaling the average odorant response to fit the records in Ringer's solution. Continuous line is a Gaussian with standard deviation of 0.14 pA.

by an enzymatic cascade similar to that used by olfactory receptor cells¹², elementary events have an average size of 1 pA in toad rods³, 0.5–1 pA in monkey rods^{13,14} and only 30 fA in monkey cones¹⁵.

Additional information on the olfactory transduction cascade is required to produce a complete quantitative explanation of our electrophysiological data. Whatever the biochemical source of the occurrence of occasionally large fluctuations at low odorant concentrations, they are likely to play a key role in determining whether a stimulus is detectable. If the olfactory neuron is exposed to an odorant for a sufficiently long time the binding of one or very few odorant molecules could produce fluctuations of a few pA that, given the high input resistance of the cell^{16–18} will initiate an action potential¹⁹. Thus high sensitivity in olfactory receptor cells could be achieved by temporal integration. □

Received 18 August; accepted 9 December 1994.

- Block, S. M. in *Sensory Transduction* (eds Coirey, D. P. & Roper, S. D.) 1–17 (Rockefeller Univ. Press, New York, 1992).
- Crawford, A. C. & Fettiplace, R. *J. Physiol.* **364**, 359–379 (1985).
- Baylor, D. A., Lamb, T. D. & Yau, K. W. *J. Physiol.* **288**, 613–634 (1979).
- Fuortes, M. G. F. & Yeandle, S. *J. gen. Physiol.* **47**, 443–463 (1964).
- Kaissling, K. E. *Rev. Neurosci.* **9**, 121–145 (1986).
- de Vries, H. L. & Stuijver, M. in *Sensory Communication* (ed. Rosenblith, W. A.) 159–167 (Wiley, New York, 1961).
- Firestein, S., Picco, C. & Menini, A. *J. Physiol.* **468**, 1–10 (1993).
- Kleene, S. J. & Gesteland, R. C. *J. Neurosci.* **11**, 3624–3629 (1991).
- Kurahashi, T. & Yau, K. W. *Nature* **363**, 71–74 (1993).
- Nakamura, T. & Gold, G. H. *Nature* **325**, 442–444 (1987).

- Zufall, F., Firestein, S. & Shepherd, G. M. *J. Neurosci.* **11**, 3573–3580 (1991).
- Lamb, T. D. & Pugh, E. N. *Trends Neurosci.* **15**, 291–298 (1992).
- Nunn, B. J. & Baylor, D. A. *Nature* **299**, 726–728 (1982).
- Baylor, D. A., Nunn, B. J. & Schnapf, J. L. *J. Physiol.* **357**, 575–607 (1984).
- Schnapf, J. L., Nunn, B. J., Meister, M. & Baylor, D. A. *J. Physiol.* **427**, 681–713 (1990).
- Trotter, D. *Pflügers Arch.* **407**, 589–595 (1986).
- Firestein, S. & Werblin, F. *Proc. natn. Acad. Sci. U.S.A.* **84**, 6292–6296 (1987).
- Kurahashi, T. *J. Physiol.* **419**, 177–192 (1989).
- Lynch, J. W. & Barry, P. H. *Biophys. J.* **55**, 755–768 (1989).

ACKNOWLEDGEMENTS. We thank V. Torre for many discussions and for critical reading of the manuscript, T. D. Lamb for suggestions and P. Gavazzo for participating in some of the experiments. This work was supported by grants from the EEC, NIH, HFSP and NATO.

Autocrine T-cell suicide mediated by APO-1/(Fas/CD95)

Jens Dhein*, Henning Walczak*,
Caroline Bäumlert†, Klaus-Michael Debatin†
& Peter H. Krammer*‡§

* Tumorimmunology Program, German Cancer Research Center,
D-69120 Heidelberg, Germany

† University Childrens Hospital, 69120 Heidelberg, Germany

§ To whom correspondence should be addressed

THE APO-1/(Fas/CD95) cell surface receptor is a member of the nerve growth factor (NGF)/tumour necrosis factor (TNF) receptor superfamily and mediates apoptosis¹⁻⁴. Peripheral activated T cells (ATC) from lymphoproliferation (lpr/lpr) mutant mice that express a reduced number of APO-1 receptors have a defect in T-cell receptor (TCR)-induced apoptosis^{5,6}. This suggests that TCR-induced apoptosis involves APO-1. We tested this hypothesis in various human T cells: (1) malignant Jurkat cells, (2) an alloreactive T-cell clone (S13), and (3) peripheral ATC. TCR triggering through immobilized anti-CD3 antibodies or *Staphylococcus enterotoxin B* (SEB) superantigen induced expression of the APO-1 ligand and apoptosis in these cells. Anti-CD3-induced apoptosis of Jurkat cells was demonstrated even in single-cell cultures. In all cases apoptosis was substantially inhibited by blocking anti-APO-1 antibody fragments and soluble APO-1 receptor decoys. The APO-1 ligand was found in the supernatant of activated Jurkat cells as a soluble cytokine. We propose that TCR-induced apoptosis in ATC can occur through an APO-1 ligand-mediated autocrine suicide. These results provide a mechanism for suppression of the immune response and for peripheral tolerance by T-cell deletion.

We compared kinetics and sensitivity to drug treatment of apoptosis induced by anti-CD3 and agonistic anti-APO-1 antibodies in Jurkat cells and in the alloreactive T-cell clone S13. Anti-APO-1 differed from anti-CD3-induced apoptosis by faster kinetics (Fig. 1a) and by lack of inhibition by cyclosporin A

(CsA), cycloheximide (CHX), and actinomycin D (Act D) (Fig. 1b, c). These differences might be explained by direct versus indirect induction of apoptosis by anti-APO-1 and anti-CD3 antibodies, respectively. We hypothesized that TCR triggering might stimulate expression of the APO-1 ligand because (1) its expression would be dependent on intact messenger RNA and protein synthesis and might, therefore, be inhibited by the above drugs, and (2) production and binding of the ligand to the APO-1 receptor might result in delayed apoptosis in contrast to direct induction of apoptosis by anti-APO-1. Expression of APO-1 ligand mRNA (2.1 kilobases (kb)) was indeed induced in anti-CD3-stimulated Jurkat cells and this induction was inhibited by CsA (Fig. 2). These results suggest that TCR-induced apoptosis involves the APO-1 ligand. This is further supported by reports demonstrating that TCR-induced apoptosis is impaired in peripheral ATC of mice with the generalized lymphoproliferative disease (gld/gld) mutation that express a non-functional APO-1 ligand^{7,11}. To demonstrate directly the involvement of the APO-1 ligand in activation-induced T-cell apoptosis we inhibited anti-CD3- or SEB-induced apoptosis in Jurkat and S13 cells, and in peripheral ATC, respectively. The following reagents were used for inhibition: (1) F(ab')₂ anti-APO-1, itself incapable of inducing apoptosis because of insufficient crosslinking of APO-1 receptors¹², and (2) soluble APO-1-Fc receptor decoys. In all cases, a substantial dose-dependent inhibition of TCR- (Fig. 3a-j) but not of dexamethasone-induced apoptosis (Fig. 3k) was found. No inhibition of TCR-induced apoptosis was observed with F(ab')₂ control antibody fragments, isotype-matched antibodies or soluble TNF receptor-Fc decoys (Fig. 3l). These data show that APO-1 and its ligand are specifically involved in TCR-induced apoptosis.

By microscopic inspection we observed that anti-CD3-induced apoptosis occurred not only in cell clusters but also in single T cells. Consistent with this observation the frequency of proliferating Jurkat cells in limiting dilution was greatly reduced in cultures treated with anti-CD3 antibodies in comparison to those in medium alone (reduction from $f=1/2$ to $f=1/9.5$; Fig. 4a). Regression analysis of the limiting dilution data revealed a

FIG. 1 Kinetics and drug inhibition of apoptosis induced by anti-APO-1 and anti-CD3 in Jurkat and human T-cell clone S13 cells. a, Kinetics of apoptosis induced by anti-APO-1 and anti-CD3 in Jurkat cells; 5×10^4 Jurkat cells (CD2⁺, CD3⁺, CD4⁺, CD18⁺ and APO-1⁺) were cultured in 0.1 ml Iscove's medium as described¹² in 96-well flat-bottom plates (Becton Dickinson) previously coated overnight with Protein A (Pharmacia) purified IgG3 anti-APO-1 ($10 \mu\text{g ml}^{-1}$) (circles) or anti-CD3 (OKT3, $100 \mu\text{g ml}^{-1}$) (triangles) antibodies. Cells were collected at the time points indicated. Apoptosis was assessed either by determination of plasma membrane integrity (filled symbols) (uptake of the DNA intercalating dye propidium iodide (PI) ($2.5 \mu\text{g ml}^{-1}$ final concentration)) or by determination of DNA fragmentation (empty symbols) (after lysing the cells in a hypotonic solution (0.1% sodium citrate, 0.1% Triton-X100) containing $50 \mu\text{g ml}^{-1}$ PI) as described²⁰ and analysed by flow cytometry. Spontaneous PI uptake in control cultures was 8% and 11%, respectively. Percentage of specific cell death was calculated as follows: $100 \times (\text{experimental PI uptake (\%)} - \text{spontaneous PI uptake of cells in medium (\%)} / (100 - \text{spontaneous PI uptake (\%)}))$. Inhibition of anti-CD3 (black bars) but not anti-APO-1 (white bars) induced apoptosis of Jurkat (b) and T cell clone S13 cells (c) by drugs. S13, an alloreactive human CD4⁺ T cell clone, was expanded as described²¹. Living cells were separated by Ficoll-Paque (Pharmacia). Jurkat or S13 cells (5×10^4) were incubated in 0.1 ml cultures in antibody-coated 96-well plates in the presence of CsA ($0.1 \mu\text{g ml}^{-1}$) or Act D ($0.1 \mu\text{g ml}^{-1}$) (b) and CsA ($0.1 \mu\text{g ml}^{-1}$) or CHX ($2 \mu\text{g ml}^{-1}$) (c). After 24 h of culture, cells were collected, stained with PI ($2.5 \mu\text{g ml}^{-1}$ final concentration) and analysed by flow cytometry. In control cultures containing drugs alone <20% of the cells were dead at 24 h. This value was regarded as spontaneous PI uptake in the formula for the calculation of specific cell death. The results shown are means of duplicate measurements. Variation was <10%. Similar results were obtained in three independent experiments.

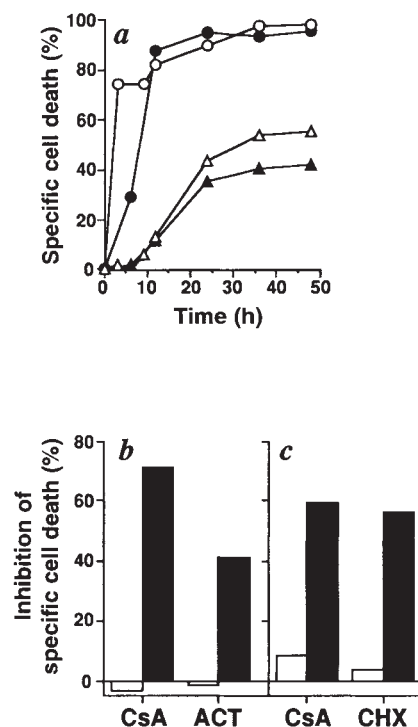
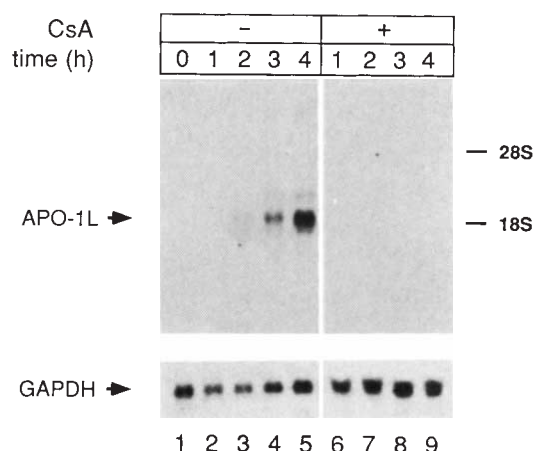


FIG. 2 APO-1 ligand (APO-1 L) mRNA expression in anti-CD3-stimulated Jurkat cells. Jurkat cells were either stimulated with coated anti-CD3 (100 $\mu\text{g ml}^{-1}$) in the absence (lanes 1–5) or presence of CsA (1 $\mu\text{g ml}^{-1}$, lanes 6–9) for the indicated times. Viability of 4-h cultures in the presence and absence of CsA was identical. Glycerolaldehyde 3-phosphate dehydrogenase (GAPDH) was used as a control for RNA loading. Increase of APO-1 ligand mRNA expression upon anti-CD3 stimulation was found in four independent experiments with similar results.

METHODS. mRNA was prepared using the Qiagen mRNA direct kit (Diagen). mRNA (3 μg per lane) was electrophoresed through a 1% agarose-formaldehyde gel and blotted to a nylon membrane. Hybridization was done under high stringency conditions with a 447 bp ^{32}P -labelled human APO-1 ligand DNA fragment derived by polymerase chain reaction (PCR) from human spleen 5'-RACE-Ready cDNA (Clontech). The oligonucleotides used for the PCR were 5'-ATAGAATTCTTAAGCTTATA-TAAGCCAAAAA as a gene-specific primer and the provided anchor primer for the first amplification, and 5'-CCTAGGTAGCTGTGGGCCCA and the anchor primer for the second amplification. The gene-specific primers are derived from regions of the rat Fas ligand corresponding to regions of high homology between TNF molecules from different species²². The amplified PCR products were subcloned into pCRII using the TA cloning kit (Stratagene). The amplified human probe used corre-



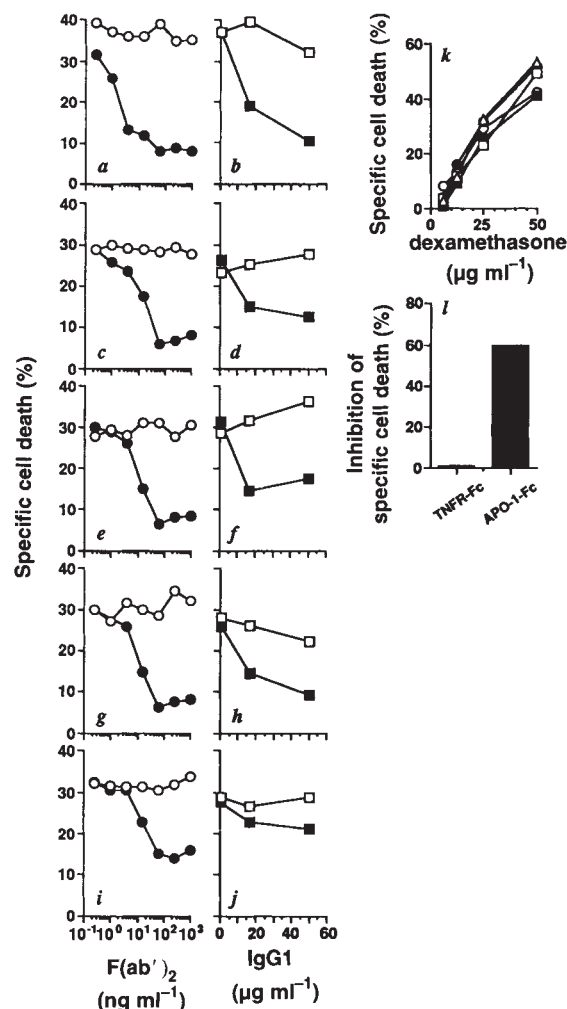
sponds to bp 355–801 of the rat Fas ligand and displays extensive homology (82%).

straight line suggesting that TCR-induced apoptosis in these cells is a single-hit event and does not necessarily require cell interactions or accessory cells. To exclude induction of apoptosis by neighbouring T cells we investigated apoptosis of single Jurkat cells plated in Terasaki microwells coated with anti-CD3. The morphology of TCR-induced apoptosis of single Jurkat cells

resembled the one of Jurkat cells in bulk culture. Because quantitative assessment of single-cell apoptosis by microscopic inspection is not objective we determined the fraction of Jurkat cells incapable of progressing from the single-cell stage. Apoptosis of cells remaining at the single-cell stage during 4 days of culture was confirmed by microscopic inspection of cellular morphol-

FIG. 3 Inhibition of TCR-induced apoptosis in T cells by F(ab')_2 anti-APO-1 and APO-1-Fc. Inhibition of anti-CD3-induced apoptosis in Jurkat cells (a, b), the T cell clone S13 (c, d), and in CD4^+ ATC (g, h), or SEB-induced apoptosis in the T-cell clone S13 (e, f) and CD4^+ ATC (i, j) by F(ab')_2 anti-APO-1 (a, c, e, g, i) and APO-1-Fc (b, d, f, h, j). Human peripheral T cells were obtained from healthy donors and isolated as described¹⁹. CD4^+ T cells were negatively selected on anti-CD8 antibody-coated magnetic beads (Dyna) and expanded for 6 to 10 days as described²⁰. Jurkat cells, S13 cells, and CD4^+ ATC were distributed in 96-well flat-bottom plates with coated anti-CD3 (100 $\mu\text{g ml}^{-1}$), soluble SEB (100 ng ml^{-1}), or medium in the presence of various concentrations of F(ab')_2 anti-APO-1 (●), F(ab')_2 control (○) derived from the isotype-matched antibody FI23¹², APO-1-Fc (■), and control IgG1 antibody I2D31D1¹² (□) (a–j). k, Lack of inhibition of dexamethasone-induced apoptosis in Jurkat cells by F(ab')_2 anti-APO-1 and APO-1-Fc. Jurkat cells were cultured in various concentrations of dexamethasone in the presence of medium alone (△), or APO-1-Fc (■), and control IgG1 antibody (□) (50 $\mu\text{g ml}^{-1}$, respectively), or F(ab')_2 anti-APO-1 (●), and F(ab')_2 control (○) (1 $\mu\text{g ml}^{-1}$, respectively). l, Inhibition of apoptosis in anti-CD3-stimulated Jurkat cells by APO-1-Fc but not by TNFR-Fc. Anti-CD3-stimulated Jurkat cells were cultured in the presence of APO-1-Fc or TNFR-Fc (50 $\mu\text{g ml}^{-1}$, respectively). Inhibition was calculated as reduction of apoptosis compared to apoptosis in the presence of control IgG1 set as 100%. Cells were cultured for 24 h (a–j, l) or 48 h (k), collected and specific cell death was assessed by measuring PI uptake as described in Fig. 1. Spontaneous PI uptake in control cultures containing proteins alone was <20%. The results shown are means of duplicate measurements. Variation was <15%. Similar results were obtained in three independent experiments.

METHODS. F(ab')_2 fragments of murine IgG3 were prepared as described¹². A cDNA encoding APO-1-Fc consisting of the extracellular domain of APO-1 linked to the Fc region of murine IgG1 through a thrombin cleavage site was constructed by PCR as described for TNFR-Fc²³. The obtained DNA fragment was digested with *HindIII* and *XbaI* and inserted into the vector pCDM8 and sequenced. Serum-free supernatants of transiently transfected COS cells were collected weekly up to 3 weeks after transfection. Murine p55 TNFR-Fc fusion protein crossreacting with human TNF α was generated in the same way. APO-1-Fc and control proteins were purified by Protein A affinity chromatography.



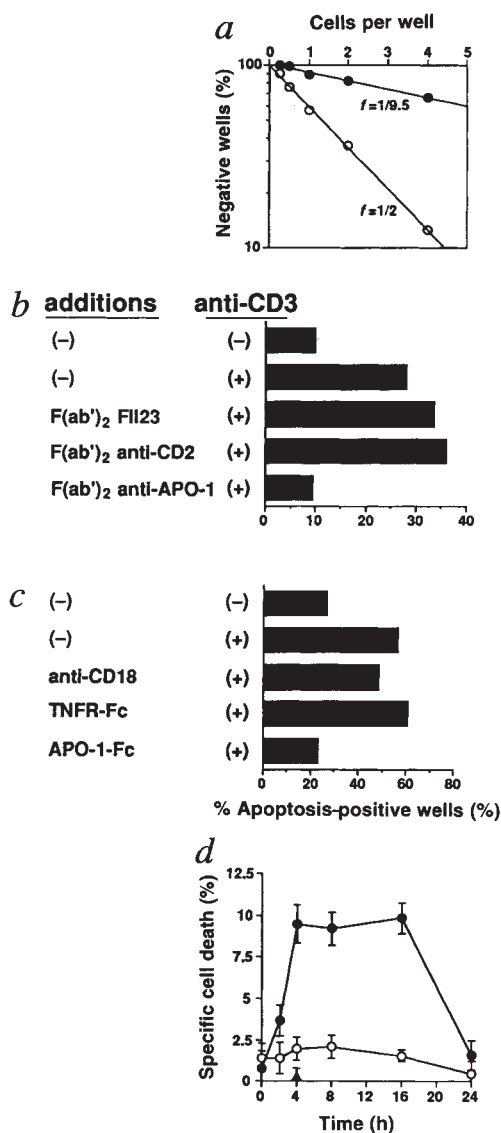


FIG. 4 Inhibition of TCR-induced soluble APO-1 ligand-mediated apoptosis in Jurkat cells by F(ab')₂ anti-APO-1 and APO-1-Fc. **a**, Regression analysis of anti-CD3-stimulated Jurkat cells in limiting dilution. Graded numbers of Jurkat cells were cultured in 96-well flat-bottom plates previously coated with anti-CD3 (○) or medium (●) as described in Fig. 1. After 12 days of culture, growth was assessed by microscopic inspection. Estimates of the frequency of proliferating cell precursors were determined by Poisson probability distribution²⁴. A repeat of the experiment gave comparable frequencies. **b**, **c**, Inhibition of TCR-induced apoptosis in single Jurkat cells by F(ab')₂ anti-APO-1 (**b**) and by APO-1-Fc (**c**). Single Jurkat cells were distributed into 60-well Terasaki microplates (Nunc, Roskilde, Denmark; 2 plates per group) using an autoclave device unit of a fluorescence activated cell sorter (FACS VANTAGE, Becton Dickinson). Terasaki plates were previously coated with anti-CD3 or medium. Cells were incubated in 10 μ l cultures in the presence of non-binding F(ab')₂ Fli23, binding F(ab')₂ anti-CD2, and F(ab')₂ anti-APO-1 (100 ng ml⁻¹), or binding anti-CD18 (murine IgG1, Immunotech, Marseille, France), TNFR-Fc, and APO-1-Fc (5 μ g ml⁻¹). Cells in coded cultures were inspected by microscope at a 24 h interval for 3 to 4 days. Wells containing a cell remaining at the single-cell stage were scored as growth negative (apoptosis positive), wells with cells progressing from the single-cell to the multiple-cell stage were scored as growth-positive wells (apoptosis negative). The number of apoptotic single cells in anti-CD3-stimulated cultures was significantly different from the one in single cell cultures in medium alone: $P=0.0029$ (**b**) and $P=0.013$ (**c**). Numbers of apoptotic cells in F(ab')₂ anti-APO-1- and APO-1-Fc-blocked cultures differed significantly from numbers in non-blocked cultures: $P=0.0021$ and $P=0.0081$, respectively. Variation was less than 20% of the mean. **d**, Soluble APO-1 ligand activity in the supernatant of anti-CD3-stimulated Jurkat. Jurkat cells (2×10^6) were incubated in 0.5-ml cultures in 24-well plates in the presence or absence of immobilized anti-CD3. At the indicated times supernatant was collected and centrifuged at 14,900g for 15 min. Anti-APO-1-sensitive (●) or -resistant (○) Jurkat cells (Jurkat^S and Jurkat^R, respectively) were resuspended in undiluted supernatant and cultured in 96-well plates (1×10^5 cells per well in 0.1 ml supernatant) for a further 24 h. Activity of the soluble APO-1 ligand in the 4 h supernatant on Jurkat^S cells was inhibited by APO-1-Fc (50 μ g ml⁻¹; ▲). The anti-APO-1 resistant variant Jurkat^R cell line was generated by continuous culture in medium with 1 μ g ml⁻¹ anti-APO-1 (IgG3, κ) for over 6 months and differed phenotypically from the parental line (see Fig. 1) only by lack of cell surface APO-1. Cell death was determined by forward/side scatter analysis as described²⁵ and calculated as in Fig. 1. Spontaneous apoptosis of Jurkat^S and Jurkat^R cells cultured in supernatant derived from unstimulated cells was 12% and 9%, respectively. s.d. was calculated from triplicate cultures. Experiments have been repeated four times with similar results.

ogy: apoptosis was evident from, for example, membrane blebbing and chromatin condensation. Our experiments showed that the fraction of single TCR-triggered Jurkat cells undergoing apoptosis was two to three times higher than in control cultures. Apoptosis of single Jurkat cells was prevented and progression from the single-cell stage was seen in the presence of blocking F(ab')₂ anti-APO-1 or soluble APO-1-Fc (Fig. 4b, c). These data suggest that TCR-induced apoptosis of a single T cell is mediated by the APO-1 receptor and its soluble ligand. This conclusion is further supported by transient inducible soluble APO-1 ligand activity found in the supernatant of anti-CD3-stimulated Jurkat cells. The soluble APO-1 ligand in the supernatant acted only on anti-APO-1 sensitive but not on anti-APO-1 resistant Jurkat target cells (Fig. 4d).

T-cell apoptosis may occur as 'fratricide' by interaction of a membrane bound form of the APO-1 ligand with the APO-1 receptor expressed on neighbouring cells^{11,13}. This mechanism might particularly be relevant for cytotoxic T lymphocyte/target cell interactions¹⁴⁻¹⁷. However, we have only detected marginal expression of the APO-1 ligand on the surface of TCR-triggered human T cells by sensitive fluorescence analysis using APO-1-Fc (data not shown). In addition, we show that TCR-induced apoptosis can occur in single T cells and can be inhibited by reagents that block APO-1 receptor/APO-1 ligand interaction. Therefore, an individual activated T cell expressed all elements

required for self-destruction: the APO-1 ligand and its receptor. A soluble form of the APO-1 ligand might result from secretion (as in the case of lymphotoxin- α) or from proteolytic cleavage (as in the case of TNF α) and might act in an autocrine or paracrine fashion. A model involving a soluble APO-1 ligand has the problem of explaining specificity and directionality of T-cell apoptosis¹⁸. Therefore, the APO-1 ligand may have a short half-life and may only act at a short distance at a sufficiently high concentration. Furthermore, a soluble APO-1 ligand may preferentially induce apoptosis in ATC that have acquired an apoptosis-sensitive phenotype¹⁹. Taken together, our results suggest a minimal model in which TCR-induced apoptosis in ATC involves autocrine soluble APO-1 ligand-mediated suicide. This mechanism may be responsible for suppression of the immune response and for peripheral tolerance by T-cell deletion. □

Received 7 September; accepted 24 November 1994.

1. Trauth, B. C. et al. *Science* **245**, 301-304 (1989).
2. Yonehara, S., Ishii, A. & Yonehara, M. *J. exp. Med.* **169**, 1747-1756 (1989).
3. Itoh, N. et al. *Cell* **66**, 233-243 (1991).
4. Oehm, A. et al. *J. biol. Chem.* **267**, 10709-10715 (1992).
5. Watanabe-Fukunaga, R., Brannan, C. I., Copeland, N. G., Jenkins, N. A. & Nagata, S. *Nature* **356**, 314-317 (1992).
6. Russell, J. H., Rush, B., Weaver, C. & Wang, R. *Proc. natn. Acad. Sci. U.S.A.* **90**, 4409-4413 (1993).
7. Russell, J. H. & Wang, R. *Eur. J. Immun.* **23**, 2379-2382 (1993).
8. Ramsdell, F. et al. *Eur. J. Immun.* **24**, 928-933 (1994).

9. Takahashi, T. et al. *Cell* **76**, 969–976 (1994).
10. Lynch, D. et al. *Immunity* **1**, 131–136 (1994).
11. Gillette-Ferguson, I. & Sidman, C. L. *Eur. J. Immun.* **24**, 1181–1185 (1994).
12. Dhein, J. et al. *J. Immun.* **149**, 3166–3173 (1992).
13. Vignaux, F. & Golstein, P. *Eur. J. Immun.* **24**, 923–927 (1994).
14. Rouvier, E., Luciani, M. F. & Golstein, P. *J. exp. Med.* **177**, 195–200 (1993).
15. Ju, S.-T., Cui, H., Panka, D. J., Ettinger, R. & Marshak-Rothstein, A. *Proc. natn. Acad. Sci. U.S.A.* **91**, 4185–4189 (1994).
16. Hanabuchi, S. et al. *Proc. natn. Acad. Sci. U.S.A.* **91**, 4930–4934 (1994).
17. Stalder, T., Hahn, S. & Erb, P. *J. Immun.* **152**, 1127–1133 (1994).
18. Lowin, B., Hahne, M., Mattmann, C. & Tschopp, J. *Nature* **370**, 650–652 (1994).
19. Klas, C., Debatin, K.-M., Jonker, R. R. & Krammer, P. H. *Int. Immun.* **5**, 625–630 (1993).
20. Nicoletti, I., Migliorati, M. C., Grignani, F. & Riccardi, C. J. *J. immun. Meth.* **139**, 271–279 (1991).
21. Kabelitz, D., Conradt, P., Schöndelmaier, S., Wagner, H. & Haas, R. *J. exp. Med.* **170**, 559–569 (1989).
22. Suda, T., Takahashi, T., Golstein, P. & Nagata, S. *Cell* **75**, 1169–1178 (1993).
23. Peppel, K., Crawford, D. & Beutler, B. *J. exp. Med.* **174**, 1483–1489 (1991).
24. Haight, F. A. *Handbook of the Poisson Distribution* (Wiley, New York, 1967).
25. Debatin, K.-M., Süß, D. & Krammer, P. H. *Eur. J. Immun.* **24**, 753–758 (1994).

ACKNOWLEDGEMENTS. J.D. and H.W. contributed equally to this work. We thank W. Müller, M. Pach, K. Hexel and A. Ehret for technical assistance, A. Benner for statistical analysis, D. Kabelitz and S. Wesselborg for S13 and J. Borst for Jurkat cells, M. Wild for providing antibodies, S. Dübel for cDNA from an IgG1 producing mouse hybridoma and H. Krafft for the TNFR-Fc plasmid.

Cell-autonomous Fas (CD95)/Fas-ligand interaction mediates activation-induced apoptosis in T-cell hybridomas

Thomas Brunner*, Rona J. Mogil*, Drake LaFace*, Nam Jin Yoo*, Artin Mahboubi*, Fernando Echeverri*, Seamus J. Martin*, Walker R. Force†, David H. Lynch‡, Carl F. Ware† & Douglas R. Green*

* Division of Cellular Immunology,

La Jolla Institute for Allergy and Immunology,

11149 North Torrey Pines Road, La Jolla, California 92037, USA

† Division of Biomedical Sciences, University of California, Riverside, California 92521, USA

‡ Department of Immunobiology,

Immunex Research and Development Corp., Seattle, Washington 98101, USA

A NUMBER of murine T-cell hybridomas undergo apoptosis within a few hours of activation by specific antigens, mitogens, antibodies against the T-cell antigen receptor, or a combination of phorbol ester and calcium ionophore^{1–3}. This phenomenon has been extensively studied as a model for clonal deletion in the immune system, in which potentially autoreactive T cells eliminate themselves by apoptosis after activation, either in the thymus⁴ or in the periphery⁵. Here we show that the Fas/CD95 receptor, which can transduce a potent apoptotic signal when ligated^{6,7}, is rapidly expressed following activation of T-cell hybridomas, as is its functional, membrane-bound ligand⁸. Interference with the ensuing Fas/Fas-ligand interaction inhibits activation-induced apoptosis. Because T-cell receptor ligation can induce apoptosis in a single T hybridoma cell, we suggest that the Fas/Fas-ligand interaction can induce cell death in a cell-autonomous manner.

To address the possible role of Fas/Fas-ligand interactions in activation-induced apoptosis in T-cell hybridomas, we examined the expression of Fas and Fas-ligand in the hybridoma line A1.1, which rapidly undergoes apoptosis following activation^{1,3}. A1.1 cells do not express cell-surface Fas unless they are activated by T-cell receptor ligation (Fig. 1a). Expression of the cell-surface Fas molecule was readily detectable at 4 h post-activation. Furthermore, expression of Fas messenger RNA after activation in these cells was optimal at ~2 h post-activation, as determined

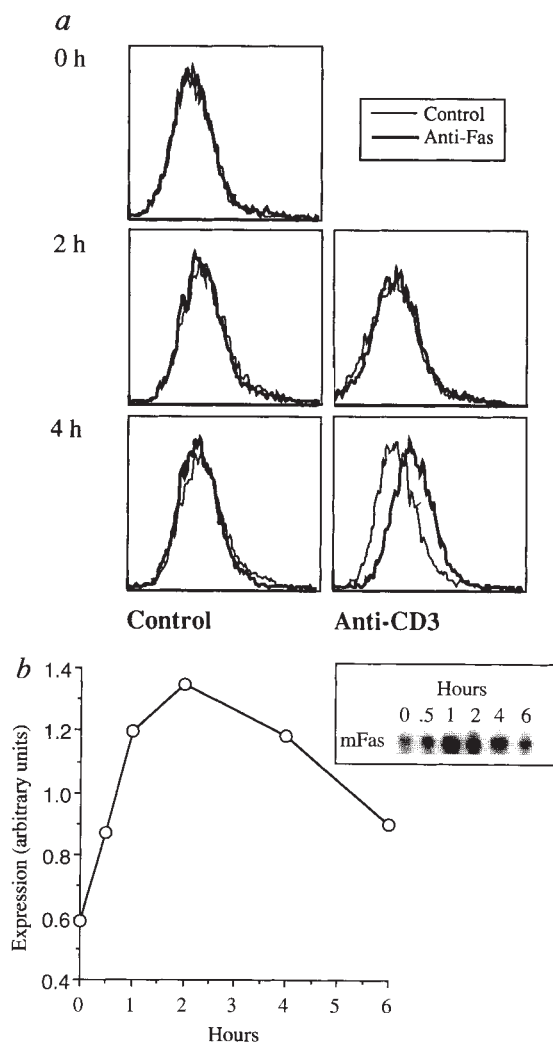


FIG. 1 Activation of T hybridoma cells induces expression of Fas. *a*, Detection of cell-surface Fas by immunofluorescence. *b*, Detection of Fas mRNA by RNase protection.

METHODS. A1.1 murine T hybridoma cells were plated (5×10^5 cells ml^{-1}) on untreated or anti-CD3 (145-2C11) treated tissue culture plastic for the indicated time. For immunofluorescence, cells were washed in PBS containing 0.2% BSA and 0.02% sodium azide (wash buffer). Cells (5×10^5 cells ml^{-1}) were incubated at room temperature in wash buffer containing $2.5 \mu\text{g ml}^{-1}$ normal mouse immunoglobulin for 15 min, after which anti-Fas antibody (Jo2; Pharmingen) was added at $2.5 \mu\text{g ml}^{-1}$ and cells were incubated for 30 min at 4°C . Cells were washed and resuspended in phycoerythrin (PE)-conjugated goat anti-hamster IgG, $2.5 \mu\text{g ml}^{-1}$ (Accurate, Westbury, NY) for 30 min at 4°C . Cells were washed once and resuspended in PBS containing 1% formaldehyde and analysed on a FACScan (Becton Dickinson). For RNase protection, labelled RNA transcripts were prepared from murine Fas and actin cDNA cloned into Bluescript and assayed using a commercial system (Ambion). The densitometer scan of the mouse Fas signal (inset) was standardized against the actin signal to determine expression (arbitrary units).

by RNase protection (Fig. 1b) and northern hybridization (data not shown).

Similarly, functional Fas-ligand is induced following activation of A1.1 cells. The cells were examined for their ability to kill a Fas⁺ target-cell line. Activation of A1.1 cells by T-cell receptor ligation (Fig. 2a) or phorbol ester and ionomycin (data not shown) induced DNA fragmentation in Fas-transfected L1210 cells⁹, but not in the Fas^{low} parental line. Similarly, activated A1.1 cells induced DNA fragmentation in a Fas⁺ P815 cell line and in Fas⁺ Jurkat cells (data not shown). Another T-

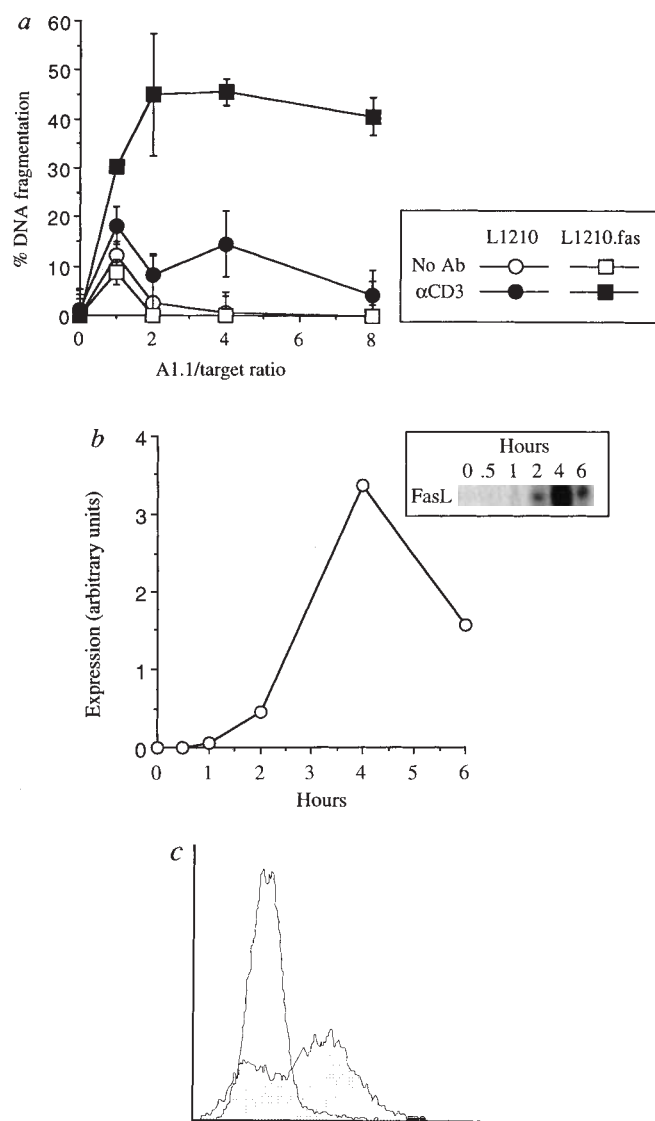


FIG. 2 Activation of T hybridoma cells induces expression of Fas-ligand. *a*, Induction of cytotoxic activity against Fas⁺ target cells. *b*, Detection of Fas-ligand mRNA by RNase protection. *c*, Detection of cell-surface Fas-ligand.

METHODS. For assessment of activation-induced cytotoxicity, Fas^{low} L1210 or Fas⁺ L1210 cells⁹ were labelled for 2 h with 5 μ Ci ml⁻¹ [³H]thymidine (79 Ci mmol⁻¹; Amersham) at 37 °C in RPMI/5% FCS. Cells were washed twice with HBSS and resuspended in RPMI/5% FCS. The target cells (2×10^4) were combined with the indicated ratio of A1.1 hybridoma cells in a total volume of 200 μ l per well in RPMI/5% FCS to untreated or anti-CD3-coated 96 well plates and incubated at 37 °C for 8 h. Unfragmented labelled high-molecular-weight DNA was collected by filtration through glass fibre filters (Pharmacia) and counted in a liquid scintillation counter²⁹. Data are expressed as per cent DNA fragmentation: $100 \times (1 - \text{c.p.m. in experimental group per c.p.m. of unstimulated targets}) \pm \text{s.e.m.}$ Each data point represents the mean of quadruplicates of a representative experiment. For RNase protection, A1.1 T hybridoma cells were activated as for Fig. 1 and analysed using labelled RNA transcripts prepared from murine Fas-ligand⁸ and actin cDNA cloned into Bluescript. The Fas-ligand signal (inset) was densitometered and standardized against the actin signal (arbitrary units, not comparable to Fig. 1). A chimaeric Fas-Fc protein was used to detect cell-surface Fas-ligand, produced as follows. A cDNA encoding the extracellular domain of Fas was isolated by PCR from the human T cell hybridoma II-23.D7. This fragment was ligated in-frame to a 710-bp cDNA fragment encoding the hinge, CH2 and CH3 domains of human IgG1 and subcloned into pCD302. For construction of the baculovirus transfer vector, a double-stranded linker was designed with an internal EcoRI site and a KpnI overhang with the sequence 5'-GAGAA-TTCGGGTAC-3' and for the bottom strand 5'-CCGAATTCTC-3'. The bottom strand was treated with T4 polynucleotide kinase before annealing the linker and subsequent ligation to the Fas-Fc KpnI/NotI fragment previously excised from pCD302. The linker added a 5'EcoRI site to the Fas-Fc fragment which allowed ligation into the EcoRI/NotI site of the baculovirus transfer vector, pVL1392 (In Vitrogen). Tn5B1-4 cells were infected with recombinant baculovirus in serum-free medium and Fas-Fc purified as described¹¹. The resulting protein resolved as a single band by SDS-PAGE. Detection of cell-surface Fas-ligand was as follows. A1.1 cells were activated on anti-CD3-coated plastic for 4.5 h, washed and stained with Fas-Fc (25 μ g ml⁻¹ for 30 min at 4 °C). Cells were washed and incubated with biotinylated rabbit anti-human IgG (Dako, Carpinteria, CA) for 30 min at 4 °C, washed, and incubated with a streptavidin-tricolour fluorescent reagent (Caltag) for 30 min at 4 °C. Cells were again washed, fixed with 1% formaldehyde in PBS and analysed using a FACScan. There was no staining in the absence of Fas-Fc (data not shown). Control (unstimulated) A1.1 cells are shown in white, activated cells in grey.

cell hybridoma, 2B4, which undergoes activation-induced apoptosis, also kills Fas⁺ targets upon activation (data not shown). In contrast, unactivated A1.1 or 2B4 cells did not kill any of the targets tested. Thus, activation of A1.1 induces functional Fas-ligand, which in turn induces apoptosis in Fas⁺ target cells. Activation-induced expression of Fas-ligand was confirmed using RNase protection, in which expression of Fas-ligand was rapidly induced following activation of A1.1 cells (Fig. 2*b*). Optimal expression was detected 4 h post-activation with anti-CD3. Activation with phorbol myristylacetate (PMA) plus ionomycin also induces expression of Fas-ligand, as detected by RNase protection (data not shown).

To detect Fas-ligand on the surface of activated T hybridoma cells, we used a fusion protein composed of human Fas and a human immunoglobulin constant region, Fas-Fc. A similar chimaeric protein was used in the purification and characterization of Fas-ligand⁸. Human Fas binds murine Fas-ligand¹⁰, so we expected this protein to detect Fas-ligand on the murine T hybridoma cells. Unactivated A1.1 T hybridoma cells did not stain with Fas-Fc, but staining was significant with cells that had been stimulated for 4.5 h with anti-CD3 (Fig. 2*c*). Similar expression was seen following activation with PMA plus ionomycin (data not shown). No staining was detected in either the unstimulated or activated cells when the secondary reagents (biotin-anti-human immunoglobulin plus fluorescent avidin) were used without Fas-Fc protein (data not shown). Thus,

although unstimulated A1.1 T hybridoma cells are negative for Fas-ligand function, mRNA or surface staining, all three are rapidly induced by activation of the cells by T-cell-receptor ligation or by phorbol ester plus ionomycin.

The expression of both Fas and Fas-ligand following activation suggested that the interaction of these molecules might be the basis for activation-induced apoptosis. To test this idea, the human Fas-Fc protein was used as a competitive inhibitor of Fas/Fas-ligand interactions¹⁰. We found that Fas-Fc effectively inhibited activation-induced apoptosis in A1.1 cells (Fig. 3), as well as in two other commonly used T cell hybridomas, DO11.10 (data not shown) and 2B4. The Fas-Fc construct inhibited the apoptotic appearance (Fig. 3*a, b*), the loss of intact nuclear DNA (Fig. 3*c*), and the DNA fragmentation associated with apoptosis (Fig. 3*d*). Although Fas-Fc blocked activation-induced apoptosis, production of interleukin-2 following activation was unaffected by the presence of the inhibitor (data not shown). In contrast to Fas-Fc, another fusion protein composed of the p60 receptor for tumour-necrosis factor (TNF) and the human immunoglobulin constant region (TNFR60-Fc)¹¹ had no effect on activation-induced apoptosis (Fig. 3*a-d*). The effects of Fas-Fc versus TNFR-Fc on activation-induced apoptosis 2B4 cells (Fig. 3*e*) was also observed in A1.1 cells (data not shown). Furthermore, a murine Fas-Fc chimaeric protein¹² likewise inhibited activation-induced apoptosis in A1.1 cells (data not shown).

In addition to apoptosis, activation of T-cell hybridomas induces inhibition of DNA synthesis and cell-cycle arrest¹³. We examined whether Fas-Fc would interfere with this activation-induced effect. As expected, DNA synthesis was inhibited by >60% in activated versus control cultures¹³ (data not shown). However, whereas Fas-Fc blocked activation-induced cell death in this experiment, DNA synthesis remained inhibited to the same extent as in the activated cultures without Fas-Fc. Thus, unlike activation-induced apoptosis, activation-induced cell-cycle arrest in T-cell hybridomas does not seem to depend on a Fas/Fas-ligand interaction.

Activation of T-cell hybridomas induces expression of both Fas and Fas-ligand on the cell surface and the subsequent interaction induces apoptosis. This explains why, when activated and non-activated T-cell hybridomas are mixed, only the activated cells die^{2,3,13}. The cells that had not been activated do not express Fas and therefore cannot receive the apoptotic signal induced by Fas-ligand. This leaves the question of whether activation-induced apoptosis is 'fratricide', in which an activated cell bearing Fas-ligand kills neighbouring activated cells bearing Fas, or 'suicide', in which these molecules can interact to induce apoptosis on a single activated cell.

We addressed this question directly by examining the fate of single cells on which the T-cell receptor was ligated. We saw no effect of cell density on the induction of apoptosis in A1.1 cells plated on anti-CD3-coated plastic (Fig. 4a). The frequency of activation-induced apoptosis among wells containing only a

single cell was not significantly different from that of wells containing cells that could potentially interact. Previous studies using larger number of cells had also suggested that activation-induced apoptosis proceeded independently of cell density¹⁴. Furthermore, we found that Fas-Fc interfered with activation-induced apoptosis in single cells (Fig. 4b). Thus, activation-induced apoptosis, proceeding through a Fas/Fas-ligand interaction, is cell autonomous and the requisite Fas/Fas-ligand interaction can proceed on a single activated cell.

It is not clear how this Fas/Fas-ligand interaction occurs on a single cell. Although production of a soluble Fas-ligand would account for this effect, we detected no Fas-ligand activity in fivefold concentrated supernatants of activated A1.1 cells using Fas⁺ L1210 cells as targets (data not shown). Alternatively, membrane-bound ligand and receptor might come into contact during movements of the cell membrane (contact of microvilli, for example), although the possibility that this occurs within the cell is unlikely because addition of competitor (Fas-Fc) blocks activation-induced cell death.

Activation-induced apoptosis in T-cell hybridomas depends on macromolecular synthesis^{2,3,14}, consistent with the need for *de novo* synthesis of Fas and Fas-ligand. It is inhibited by cyclosporin A^{1,15}, glucocorticoids^{16,17} and retinoids^{18,19}, which might inhibit expression of the interacting molecule(s) after activation. Activation-induced apoptosis also depends on the Myc/Max heterodimer^{20,21} and the Nur77 steroid receptor^{22,23}, which are transcriptional regulators that might control expression of Fas

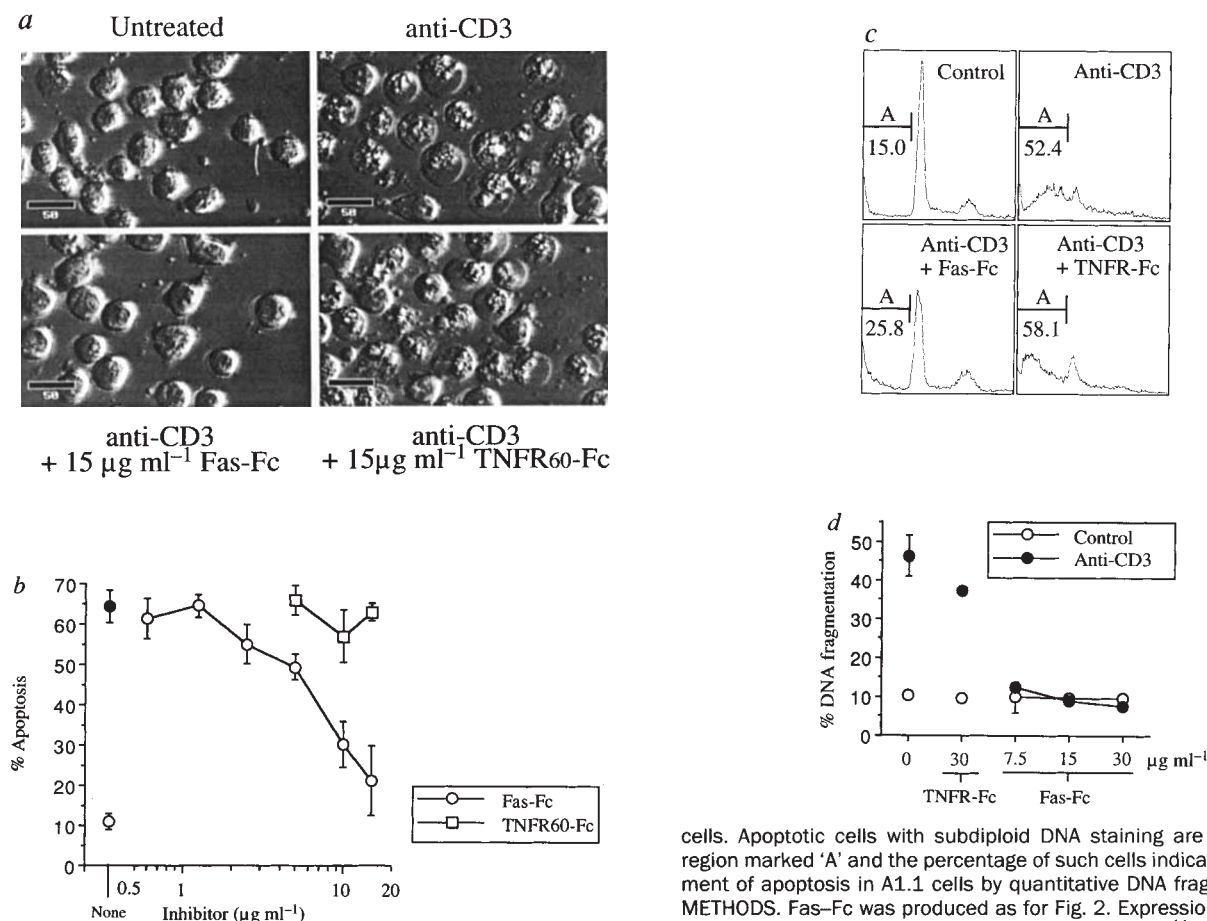


FIG. 3 Activation-induced apoptosis in T hybridoma cells is inhibited by a Fas-Fc chimaeric protein. *a*, Morphological appearance of A1.1 cells activated on anti-CD3-coated plastic in the presence or absence of Fas-Fc (15 µg ml⁻¹) or TNFR60-Fc¹¹ (15 µg ml⁻¹). The scale bar depicts arbitrary units. *b*, Assessment of apoptosis in A1.1 cells by acridine orange/ethidium bromide fluorescence staining. *c*, Assessment of apoptosis in 2B4 cells by propidium iodide staining of permeabilized

cells. Apoptotic cells with subdiploid DNA staining are shown in the region marked 'A' and the percentage of such cells indicated. *d*, Assessment of apoptosis in A1.1 cells by quantitative DNA fragmentation. METHODS. Fas-Fc was produced as for Fig. 2. Expression and purification of TNFR60-Fc in baculovirus has been described¹¹. For morphological assessment of apoptosis, cells were photographed directly (*a*) or examined under ultraviolet light after staining with acridine orange/ethidium bromide³⁰. Apoptotic cells showed extensive membrane blebbing, condensed chromatin and nuclear fragmentation. Permeabilized 2B4 cells were stained with propidium iodide for cell-cycle analysis³⁰. A1.1 cells labelled 16 h earlier with ³H-thymidine were used for quantitative DNA fragmentation^{20,30}.

and/or Fas-ligand following activation. Many observations on activation-induced apoptosis in T-cell hybridomas need to be re-evaluated in terms of Fas and Fas-ligand expression.

Activation of proteases has been implicated in activation-induced apoptosis²⁴ and we have found that an intracellular protease substrate is cleaved rapidly during both activation- and Fas-induced apoptosis (S.J.M. and D.R.G., unpublished), suggesting that proteases act downstream of Fas ligation. Protease activation of cyclin-dependent kinases is required for at least one form of apoptosis²⁵ and activation-induced apoptosis also depends upon such kinases²⁶. Thus, Fas ligation, protease activation, and cyclin-dependent kinase function might act in series to mediate activation-induced apoptosis.

A system has been described in which stimulation of previously activated and rested T cells induces apoptosis²⁷, an effect that depends on functional Fas and Fas-ligand as it is not evident in T cells from Fas-defective *lpr* mice; neither is peripheral deletion of T cells following antigen stimulation *in vivo*²⁸ (R.J.M. *et al.*, submitted), so this process might also involve an activation-induced Fas/Fas-ligand interaction. Using a previously described method¹², we observed that Fas-Fc blocks apoptosis induced by activation of the Th1-type CD4⁺ T cell line RA-8: apoptosis was >60% after 18 h culture on anti-TCR β -coated plastic (~10% in uncoated plates), and this activation-induced

apoptosis was reduced to ~24% in the presence of Fas-Fc (data not shown). Thus, in addition to its role in T-cell hybridomas, a Fas/Fas-ligand interaction seems to mediate activation-induced apoptosis in cases involving non-transformed T cells, supporting a possible immunoregulatory role for apoptosis induced by Fas during normal immune responses. □

Received 26 September; accepted 9 December 1994.

- Shi, Y., Sahai, B. M. & Green, D. R. *Nature* **339**, 625–626 (1989).
- Ucker, D. S., Ashwell, J. D. & Nickas, G. J. *Exp. Med.* **169**, 3461–3469 (1989).
- Shi, Y. *et al.* *J. Immun.* **144**, 3326–3333 (1990).
- Smith, C. A., Williams, G. T., Kingston, R., Jenkinson, E. J. & Owen, J. J. T. *Nature* **337**, 181–184 (1989).
- Kawabe, Y. & Ochi, A. *Nature* **349**, 245–248 (1991).
- Yonehara, S., Ashii, A. & Yonehara, M. *J. exp. Med.* **169**, 1747–1756 (1989).
- Trauth, B. C. *et al.* *Science* **245**, 301–305 (1989).
- Suda, T., Takahashi, T., Golstein, P. & Nagata, S. *Cell* **75**, 1169–1178 (1993).
- Rouvier, E., Luciani, M. F. & Golstein, P. *J. exp. Med.* **177**, 195–200 (1993).
- Ramsdell, F. *et al.* *Eur. J. Immun.* **24**, 928–933 (1994).
- Crowe, P. D., VanArsdale, T. L., Walter, B. N., Dahms, K. M. & Ware, C. F. *J. Immun. Meth.* **168**, 79–89 (1994).
- Ramsdell, F. *et al.* *Int. Immun.* **6**, 1545–1553 (1994).
- Ashwell, J. D., Cunningham, R. E., Noguchi, P. D. & Hernandez, D. J. *exp. Med.* **165**, 173–194 (1987).
- Green, D. R., Bissonnette, R. P., Glynn, J. M. & Shi, Y. *Sem. Immun.* **4**, 379–388 (1992).
- Mercep, M., Noguchi, P. D. & Ashwell, J. D. *J. Immun.* **142**, 4085–4092 (1989).
- Zacharchuk, C. M., Mercep, M., Chakroborty, P. K., Simons, S. S. & Ashwell, J. D. *J. Immun.* **145**, 4037–4045 (1990).
- Watanabe, M., Hanaoka, S. & Sato, K. *Eur. J. Immun.* **21**, 643–648 (1991).
- Iwata, M., Mukai, M., Nakai, Y. & Iseki, R. *J. Immun.* **149**, 3302–3308 (1991).
- Yang, Y., Vacchio, M. S. & Ashwell, J. D. *Proc. natn. Acad. Sci. U.S.A.* **90**, 6170–6174 (1993).
- Shi, Y. *et al.* *Science* **257**, 212–214 (1992).
- Bissonnette, R. P., McGahon, A. & Green, D. R. *J. exp. Med.* **180**, 2413–2418 (1994).
- Liu, Z.-G., Smith, S. W., McLaughlin, K. A., Schwartz, L. M. & Osborne, B. A. *Nature* **367**, 281–284 (1994).
- Woronicz, J. D., Calnan, B., Ngo, V. & Winoto, A. *Nature* **367**, 277–281 (1994).
- Sarin, A., Adams, D. H. & Henkart, P. A. *J. exp. Med.* **178**, 1693–1700 (1993).
- Shi, L. *et al.* *Science* **263**, 1143–1145 (1994).
- Fotadar, R. *et al.* *Molec. cell. Biol.* (in the press).
- Russell, J. H., Rush, B., Weaver, C. & Wang, R. *Proc. natn. Acad. Sci. U.S.A.* **88**, 4409–4413 (1993).
- Singer, G. G. & Abbas, A. K. *Immunity* **1**, 365–371 (1994).
- Matzinger, P. *J. Immun. Meth.* **145**, 185–192 (1991).
- McGahon, A. J. *et al.* in *Cell Death* (eds Schwartz, L. M. & Osborne, B. A.) (Academic, Orlando, in the press).

ACKNOWLEDGEMENTS. T.B. and R.J.M. contributed equally to this work. Supported by grants from the US NIH; S.J.M. is a Wellcome Trust Fellow and T.B. is a fellow of the Swiss NSF. We thank W. Nishioka, A. McGahon, F. Ramsdell and A. Fotadar for discussion, P. Crowe and B. Walter for purification of TNFR-Fc and Fas-Fc, and S. Nagata and P. Golstein for reagents.

Fas(CD95)/FasL interactions required for programmed cell death after T-cell activation

Shyr-Te Ju^{*†}, David J. Panka[‡], Hailli Cui^{*}, Rachel Ettinger[†], Maan El-Khatib^{*}, David H. Sherr^{†§}, Ben Z. Stanger^{||} & Ann Marshak-Rothstein^{†§}

^{*} Arthritis Center, and Departments of [†] Pathology and Laboratory Medicine, [‡] Microbiology, and [§] School of Public Health, Boston University School of Medicine, 80 East Concord Street, Boston, Massachusetts 02118, USA

^{||} Department of Genetics, Harvard Medical School, 200 Longwood Avenue, Boston, Massachusetts 02115, USA

RECEPTOR crosslinking of T-cell hybridomas induces cell activation followed by apoptosis^{1–6}. This activation-induced cell death requires *de novo* synthesis of RNA and proteins^{1–3}, but the actual gene products that provide the death signal have not been identified^{4–6}. We show here that receptor crosslinking induces Fas ligand and upregulates Fas, and that the ensuing engagement of Fas by Fas ligand activates the cell-death programme. Cell death, but not activation, can be selectively prevented by a soluble Fas-immunoglobulin fusion protein. Thus, Fas and Fas ligand are the

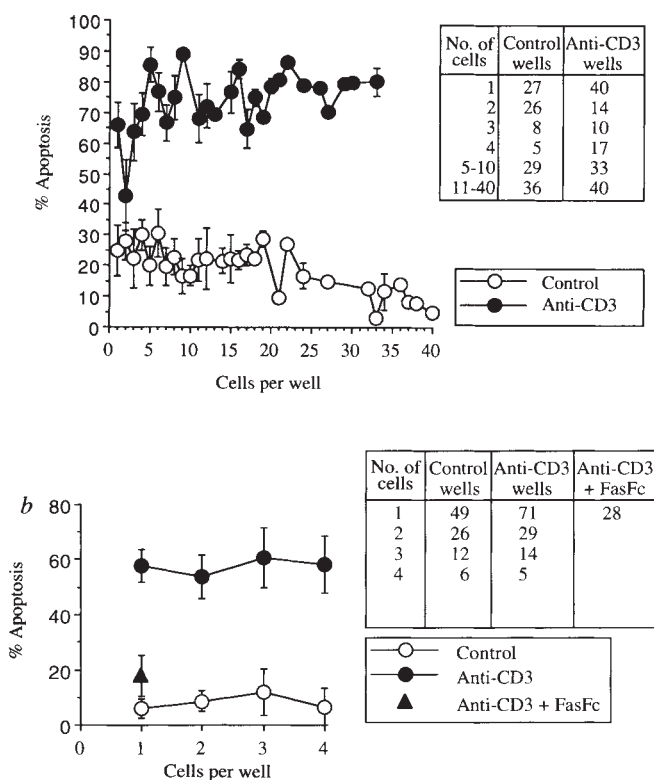


FIG. 4 Activation-induced apoptosis in T hybridoma cells is not density-dependent. *a*, Activation-induced apoptosis in A1.1 cells at a range of cell densities. *b*, Inhibition by Fas-Fc of activation-induced apoptosis in single A1.1 cells.

METHODS. A1.1 T hybridoma cells were plated in limiting numbers ranging from 1–40 cells per well in 10 μ l in untreated or anti-CD3-coated tissue culture microwells. Cells were cultured overnight, stained with acridine orange/ethidium bromide³⁰ *in situ* and viewed under ultraviolet on an inverted microscope. The total numbers of normal and apoptotic cells in each well were counted: the number of wells with a given number of cells is shown for each condition (inset). For inhibition with Fas-Fc, cells were plated in microwells at ~1 cell per well and counted under the microscope; Fas-Fc was added (25 μ g ml⁻¹) to some anti-CD3-coated wells containing single cells. Wells were checked for apoptosis after culture.

death-gene products, and their interaction accounts for the molecular mechanism of activation-induced T-cell death.

A death programme that can be coupled to T-cell activation (activation-induced cell death, AICD) occurs within individual cells¹⁻³. We tested whether the T-cell hybridomas K31H28 and ME1 met this criterion before we used them to investigate the molecular mechanism of AICD⁸. Cell death was extensive when these hybridoma T cells were cultured for 12 h or more in anti-CD3-coated plates (Fig. 1a); CD3⁻ BW5147 cells did not die under the same conditions. AICD of the V β 8⁺ K31H28 cells but not the V β 4⁺ ME1 cells could be induced in anti-V β 8-coated plates (Fig. 1b). Co-culture of the V β 8⁺ K31H28 cells with the V β 4⁺ ME1 cells in anti-V β 8-coated plates induced extensive death of K31H28 cells and a moderate level of death of bystander ME1 cells. Similar results were obtained in the reciprocal experiments using anti-V β 4-coated plates (Fig. 1b). Plate-bound anti-CD3 also induced hybridoma cell death under single-cell culture conditions. Hybridoma cells were seeded at one cell per well in anti-CD3-coated plates together with irradiated feeder cells. The frequency of clone formation 11 days later was dramatically reduced in anti-CD3 plates when compared with controls. The frequency of clone formation by BW5147 cells was not diminished in the anti-CD3-coated plates (Fig. 1c). Our data show that AICD of hybridoma T cells can be a cell-autonomous process¹⁻³, but bystander lysis also occurs.

If the Fas/FasL (Fas ligand) death program is coupled with T-cell activation, activated T cells must co-express Fas/FasL. The expression of Fas/FasL messenger RNA was examined by reverse-transcription polymerase chain reaction (RT-PCR) and RNase protection (Fig. 2). Dexamethasone, a glucocorticoid known to induce cell death by a mechanism distinct from anti-CD3, was included for comparison^{2,8}. Both K31H28 and ME1 constitutively expressed a moderate level of Fas mRNA, whereas FasL was undetectable. Fas mRNA levels were increased after activation with anti-CD3, and FasL mRNA was dramatically upregulated. By contrast, dexamethasone failed to induce FasL

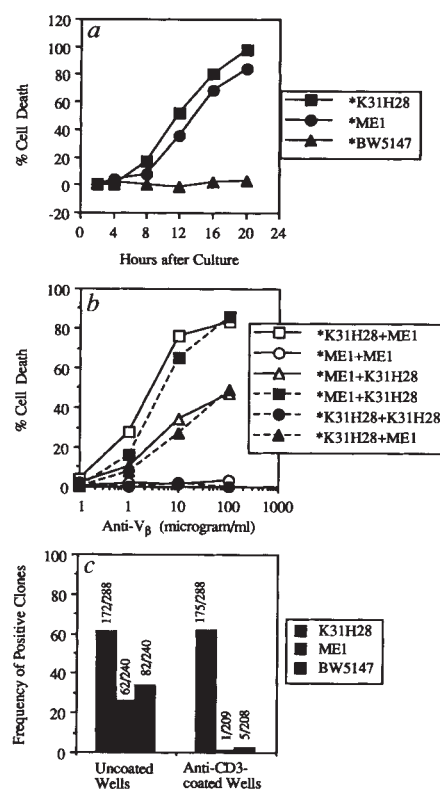
mRNA and downregulated Fas mRNA even though extensive cell death was induced. Anti-CD3 did not induce BW5147 cells to express FasL mRNA or upregulate Fas mRNA levels (Fig. 2b).

Cell-surface expression of Fas/FasL was assessed by functional analysis. These hybridomas express Fas, as shown by their susceptibility to the Fas-dependent cytotoxicity of cloned CD4⁺ Th1 cells^{9,10} (Fig. 3a). An affinity-purified Fas-immunoglobulin fusion protein (Fas-Ig) inhibited the lysis of hybridoma targets in a dose-dependent fashion. Under identical conditions, both the unbound fraction and a control p60TNFR-Ig fusion protein were not inhibitory (Fig. 3b). The functional expression of FasL by T-cell hybridomas was assessed using ⁵¹Cr-labelled P815 as the target population and anti-CD3 monoclonal antibody as the activation agent^{11,12}. Strong lysis of the target cells was observed after 16 h incubation (Fig. 3c). Killing was not detected in the absence of anti-CD3 antibody or using CD3⁻ BW5147 cells as effectors. Affinity-purified Fas-Ig inhibited the hybridoma effector activity whereas both the unbound fraction and the control p60TNFR-Ig did not (Fig. 3d). Together, these results indicate that activated hybridoma T cells co-express functional Fas and FasL proteins.

AICD could also be inhibited completely in a dose-dependent manner by Fas-Ig but not by control p60TNFR-Ig (Fig. 4a, b). Unlike AICD, interleukin-2 (IL-2) production was not affected, indicating that Fas-Ig did not interfere nonspecifically with activation dependent on the T-cell receptor (TCR) (Fig. 4c, d). Fas-Ig failed to inhibit dexamethasone-induced death of hybridoma T cells, anti-IgM-induced death of WEHI-231 B cells, and the death of HT-2 cells deprived of IL-2 (not shown). Thus, the Fas/FasL-mediated death programme is selectively coupled with TCR-dependent activation.

De novo gene transcription is necessary for AICD, and we have identified the FasL gene as one of these genes. The immediate-early gene products necessary for AICD may be the intermediates that couple T-cell activation with FasL induction⁴⁻⁶. For

FIG. 1 AICD of T-cell hybridomas. In a, plates (96-well) were coated with purified rabbit anti-hamster Ig antibodies (10 μ g per well per 100 μ l in 50 mM bicarbonate buffer, pH 9) at 37 $^{\circ}$ C for 1 d, washed, and then coated with 100 μ l of 145-2C11 anti-CD3 culture supernatants. Plates were washed 3 times with medium before use. T-cell hybridomas K31H28 and ME1, and fusion partner BW5147 cells were ⁵¹Cr-labelled and tested at 3×10^4 per well for AICD as described⁸. Cells cultured in uncoated wells were used to determine background release. The percentage cell death was determined at various times after culture by the formula: (c.p.m. released from the experimental group - c.p.m. of background release)/(c.p.m. released by 0.5% Triton X-100 - c.p.m. of background release). Background release increased with time and was <12% for 4 h assays to <36% for 20 h assays. In b, wells coated with various concentrations of purified KJ16 anti-V β 8 (solid lines) or KT4 anti-V β 4 (dashed lines) antibodies were used to induce AICD of hybridoma targets *K31H28 (V β 8⁺) and *ME1 (V β 4⁺). Labelled (*) cells (10^4 per well) were added with unlabelled cells (10^4 per well) to assess autologous and bystander lysis. The percentage cell death was determined at 14 h after incubation and background lysis was <25%. In c, hybridoma and BW5147 cells were cloned by limiting dilution at 1 cell per well in uncoated wells and in anti-CD3-coated wells. Feeder cells (10^5 per well) from 2-week-old *gld* spleens and irradiated at 3,000R were added 2 h later. Clone formation was determined by microscopic examination at 11 d after culture. Total numbers of positive clones and of wells are shown and were used to calculate the frequency of positive clones.



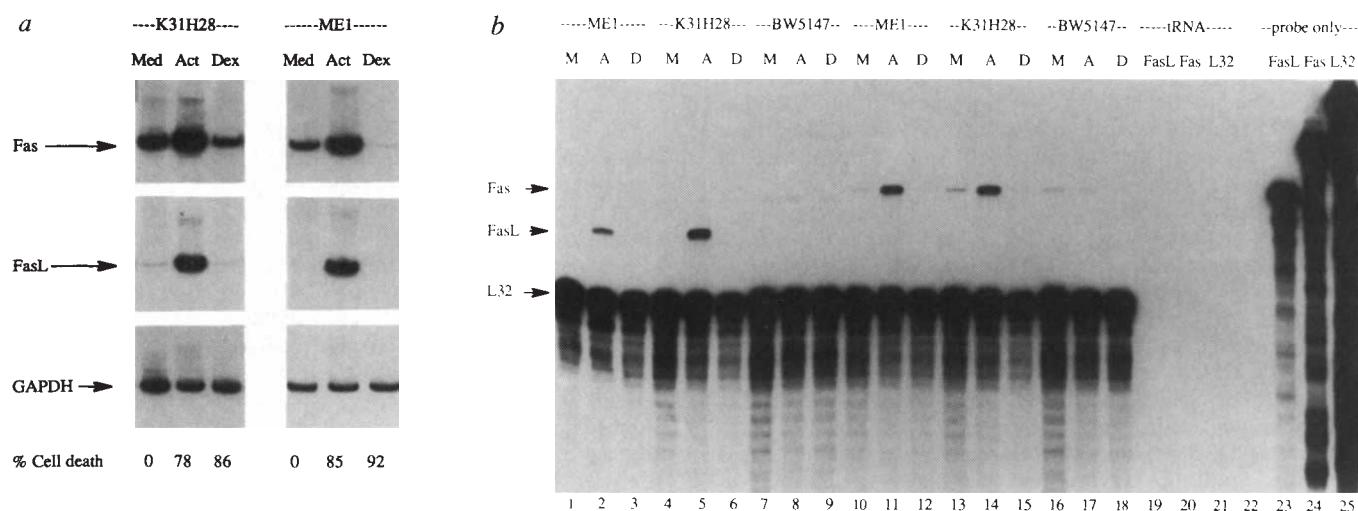


FIG. 2 Co-expression of Fas and FasL mRNA by anti-CD3-activated hybridoma T cells. For **a**, plates (24-well) were coated essentially as for Fig. 1, using 0.5 ml affinity-purified rabbit anti-hamster Ig antibodies followed by 1 ml anti-CD3 culture supernatant. K31H28 and ME1 hybridoma cells (3×10^5 per well) were cultured in uncoated plates (Med) or in plates coated with anti-CD3-mAb (Act). Hybridoma cells were also cultured in uncoated wells in the presence of 5 μ M dexamethasone (Dex). After incubation for 12 h, cells were collected and RNA extracts were prepared²³. The mRNA levels of Fas, FasL and GAPDH (glyceraldehyde-3-phosphate dehydrogenase) were assessed by RT-PCR. A similar experiment was carried out with labelled hybridoma cells in 96-well plates and the per cent cell death was determined after incubation for 16 h. In **b**, the mRNA levels of FasL (lanes 1–9) and Fas (lanes 10–18) were further assessed by RNase protection assay using total RNA obtained from hybridomas and control BW5147 cells that had been cultured in the presence of medium (M), plate-bound anti-CD3 (A), or dexamethasone (D). Protected fragments corresponding to FasL, Fas and L32 RNAs are indicated by arrows. A faint band corresponding to full-length undigested FasL probe can be seen. Lanes 19–21, tRNA controls; lane 22, empty; and lanes 23–25 riboprobes alone. **METHODS.** Total RNA was prepared using the RNAzol-B method²³. Single-stranded cDNA was prepared from RNA in a 20- μ l reaction mixture containing 45 pmol oligo(dT)₁₈, and 200 units of murine MLV reverse transcriptase; 2 μ l of the reaction mixture was diluted with 18 μ l PCR buffer containing 1.3 mM MgCl₂, 20 pmol of forward and reverse oligonucleotides, and 1 μ Ci of [³²P]dCTP (3,000 μ Ci per mmol; NEN). Three PCR reactions were set up for each cDNA, corresponding to Fas (forward: 5'-ATCCGAGCTCTGAGGAGCGGGTTCATGAAAC-3'; reverse: 5'-GGAGGTTCTAGATTCAGGTCATCCTG-3') (ref. 24); FasL (forward: 5'-CAGCTCTCCACCTGCAGAAGG-3'; reverse: 5'-AGATTCCTCAAAATTGATC-

AGAGAGAG-3') (ref. 25); and GAPDH (forward: 5'-CCATCACC-ATCTCCAGGAG-3', reverse: 5'-CCTGCTTACCACCTTCTTG-3') (ref. 26). The products were resolved on a 4% non-denaturing polyacrylamide gel followed by autoradiography. Plasmids containing the full-length cloned Fas and FasL genes (sequence verified) were used as controls for size and specificity of the PCR (not shown). RNase protection analysis: Plasmids containing the 3' coding portion of the Fas cDNA cloned into pBluescriptKS (Stratagene Cloning Systems) and the full coding region of the FasL cDNA cloned into pGEM-1 (Promega) were used as templates to generate antisense probes. Templates were linearized at the XbaI site of Fas or the AclI site of FasL to yield protected fragments with predicted molecular sizes of approximately 388 nt and 360 nt, respectively. Riboprobes were synthesized using an *in vitro* transcription kit (Stratagene), with the addition of 20 μ M cold rUTP and 100 μ Ci ³²P-UTP (NEN). After removal of the DNA template with DNase I, probes were purified using an ultrafree MXC 30,000 retention filter (Millipore). As an internal control for RNA quantity, a ribosomal L32 probe recognizing 279 bases of the processed pseudogene 4A was synthesized at 10% the specific activity of other probes and included in all samples as described²⁷. Between $1-2 \times 10^5$ c.p.m. of the Fas and FasL probes and 2×10^4 c.p.m. of the L32 probe were mixed with 10 μ g of the hybridoma RNAs or 20 μ g tRNA controls and allowed to hybridize in PIPES buffer with 50% formamide overnight at 50 °C. Unannealed RNA was digested with RNase A (40 μ g ml⁻¹) and RNase T1 (1.3 μ g ml⁻¹) at room temperature for 20 min. RNase was inactivated with proteinase K and SDS. Samples were extracted with phenol/chloroform, ethanol-precipitated, washed, and boiled for 3 min before loading on a 5% polyacrylamide gel. After electrophoresis, the gel was dried and exposed at -70 °C with an intensifying screen.

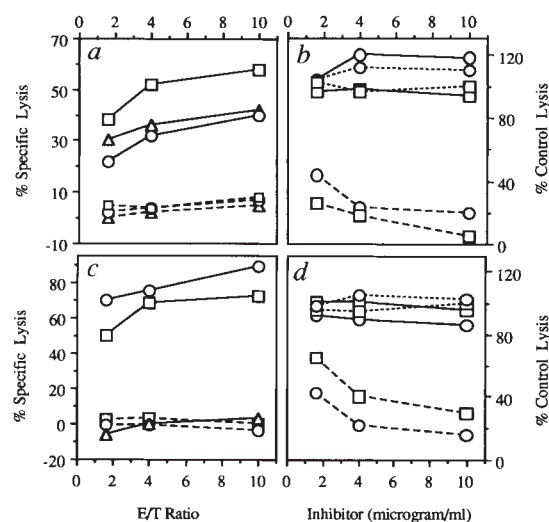
example, Nur77 is itself a transcription factor, and may act as a positive regulator of FasL transcription. Further studies of FasL gene regulation should refine the framework of the molecular mechanism of AICD established here.

Inasmuch as AICD is critically involved in clonal deletion, our study suggests that potent activation through high-affinity interactions between TCR and major histocompatibility complex (MHC) peptide can result in strong Fas/FasL co-expression. The death programme is then generated by Fas/FasL interactions on the surface of an activated cell or between conjugated cells. Because the mechanism requires the participation of two death-gene products (two genes are needed for one death signal), a defect in either one will impair AICD. The inability of *lpr* and *gld* peripheral T cells to undergo AICD strongly supports the critical role of Fas/FasL molecular interactions in peripheral T-cell tolerance^{13,14}.

Whether Fas/FasL interactions participate in thymocyte negative selection remains unresolved¹⁵⁻¹⁸. Our model provides a new test ground to determine the role of Fas/FasL in this process. Moreover, cyclosporin A and 9-*cis*-retinoic acid are known inhibitors of AICD^{5,3,19}, and both block thymic negative selection^{19,21}. Both reagents also inhibit FasL induction in T-cell hybridomas (H.C. *et al.*, manuscript submitted). Thus, signals that control thymic negative selection may act by modulating Fas/FasL expression. Based on similar considerations, E64 (a cysteine protease inhibitor) has been shown to inhibit AICD of hybridoma T cells and T cells from HIV⁺ patients, raising the possibility that Fas/FasL interactions are responsible for patients' T-cell loss²². The ability of Fas-Ig to inhibit thymic negative selection and T-cell death in HIV⁺ patients will directly determine the role of Fas/FasL interactions in these two important immune phenomena. □

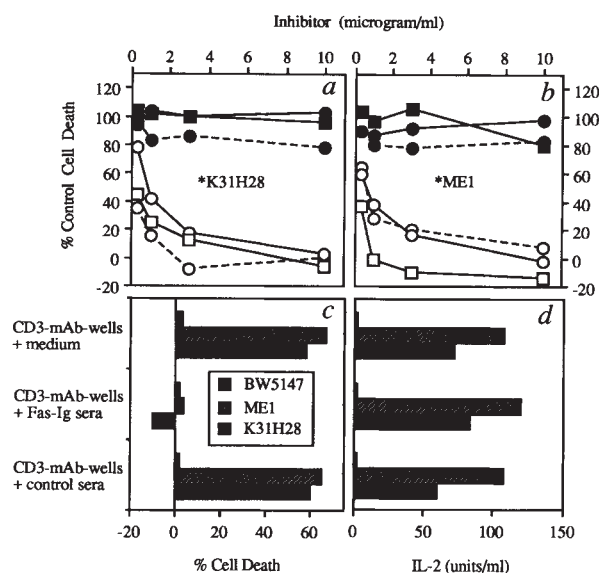
FIG. 3 Functional expression of Fas and FasL by T-cell hybridomas. In a, hybridomas, K31H28 (○), ME1 (□), and control BW5147 cells (△) were labelled with ^{51}Cr and used as targets (2×10^4 per well) for the CD4 $^{+}$ Th1 clone C7 effectors in a 5-h lectin-dependent cytotoxic assay as previously described (solid lines)¹⁰. Controls for nonspecific lysis were determined in the absence of concanavalin A (dashed lines). The per cent specific lysis was calculated by the formula: $100 \times (\text{experimental release} - \text{background release}) / (\text{total release} - \text{background release})$. Background release (4–19%) was obtained from targets cultured in medium alone. Total release was determined by lysing targets with 0.5% Triton X-100. In b, lysis of targets was assayed at an E/T ratio of 3 in the presence of increasing concentrations of affinity-purified Fas-Ig fusion protein (dashed lines). Control inhibitors included equal amounts of human Ig added into the unbound fraction (solid lines), and purified p60TNFR-Ig (dotted lines). The results were expressed as percentage of control lysis by the formula: (% specific lysis in the presence of inhibitor / % specific lysis in the absence of inhibitor) $\times 100$. Control lysis determined in the absence of inhibitor were 28 and 35% for K31H28 and ME1, respectively. In c, the per cent specific lysis of P815 target cells by hybridoma effectors at various E/T ratio was determined after incubation for 16 h. Cell mixtures were cultured either in the presence (solid lines) or absence (dashed lines) of anti-CD3 mAb (25% of 145-2C11 culture supernatant) as previously described¹². Background release for the P815 cells was 19%. In d, the per cent specific lysis of P815 target cells by T-cell hybridomas at an E/T ratio of 1 was determined after incubation for 16 h in the presence of affinity-purified Fas-Ig fusion protein (dashed lines). Control inhibitors were the unbound fraction reconstituted with human Ig (solid lines), or purified p60TNFR-Ig (dotted lines). The results were expressed as per cent of control lysis. Control lysis determined in the absence of inhibitor were 64% for K31H27 and 67% for ME1 cells. All experiments were carried out in duplicate with less than 5% variation.

METHODS. Construction and synthesis of Fas-Ig fusion protein. A cDNA construct consisting of the extracellular domain of mouse Fas joined to the hinge and constant region of human IgG1 was made by cloning the PCR-amplified Fas extracellular domain, including the leader sequence, into a human IgG1 expression vector digested with *XhoI* and *BamHI*. The oligonucleotides used to amplify the Fas extracellular domain from nucleotides 27 to 557 in the published mouse cDNA sequence were 5'-CCCTCGAGCCGCTGTTTCCCTTGCTGCA and 5'-CCCGGATCCCCGC-GATTCTGGGACTTTGTTCC-3'. To make stable lines expressing Fas-Ig, a *HindIII*/*BglII* partial fragment consisting of the Fas-Ig fusion cDNA was cloned into the *HindIII* and *BamHI* sites of the APTag-1 expression



vector²⁸. NIH3T3 cells were transfected with 20 μg Fas-Ig APTag, linearized with *ClaI*, and 1 μg pSV7neo, linearized with *EcoRI*, by calcium phosphate precipitation. Neomycin-resistant colonies were selected with 0.4 mg ml^{-1} G418 and tested for expression of Fas-Ig fusion protein by using an anti-IgG sandwich ELISA. Preparation of Fas-Ig from Fas-Ig-3T3 transfectants: culture supernatants were concentrated by 50% saturated ammonium sulphate precipitation. After dialysis against DMEM medium, aliquots were passed through a protein A column. Bound materials were eluted with 0.2 M glycine-HCl buffer, pH 2.5, and immediately neutralized with a 2 M Tris-HCl buffer, pH 8. Both the unbound fraction and the eluted fraction were dialysed against DMEM. The concentration of Fas-Ig in each preparation was determined by ELISA using human Ig as standard. The same amount of human Ig was added to the unbound fraction and used as a control. The p60TNFR-Ig, affinity-purified on a protein A column, was also used as a control²⁹.

FIG. 4 Fas-Ig fusion proteins inhibit AICD of T cell hybridomas but not IL-2 production. ^{51}Cr -labelled K31H28 (a) or ME1 cells (b) were cultured at 3×10^4 cells per well for 16 h in anti-CD3-coated or uncoated control wells (solid lines). Bystander lysis of hybridomas was established as described for Fig. 1b and carried out in wells coated with 50 $\mu\text{g ml}^{-1}$ anti- $\nu\beta$ antibodies (dashed lines). Cells were incubated in the presence or absence of various amounts of inhibitors. The inhibitors were: sera (0.4 mg ml^{-1} Fas-Ig) from SCID mice previously inoculated with Fas-Ig 3T3 transfectants (○), unbound fraction of Fas-Ig-containing sera after adsorption on a protein A column (■), affinity-purified p60TNFR-Ig (●), and affinity-purified Fas-Ig (□). The per cent cell death was determined as for Fig. 1. For a and b, background release was 35% for K31H28 and 26% for ME1. In the absence of inhibitor, the percentage cell death was 75% for K31H28 and 64% for ME1, and the bystander cell death was 42% for K31H28 and 44% for ME1. The results were expressed as per cent control cell death by the formula: % cell death in the presence of inhibitor / % cell death in the absence of inhibitor. For c and d, normal and Fas-Ig-containing sera (1 μl) were used to inhibit the AICD of K31H28 and ME1 hybridoma T cells in anti-CD3 coated wells. The CD3 $^{-}$ BW5147 line was included as control. Background release was 30% for K31H28, 26% for ME1 and 20% for BW5147. The 16-h culture supernatants were assayed for IL-2 levels as described³⁰. IL-2 activity was not detected in supernatants from cells cultured in uncoated plates. Similar results were obtained with affinity-purified Fas-Ig and the control p60TNFR-Ig fusion proteins.



Received 27 July; accepted 22 December 1994.

1. Ashwell, J. D., Cunningham, R. E., Noguchi, P. D. & Hernandez, D. J. *exp. Med.* **165**, 173–194 (1987).
2. Shi, Y., Sahai, B. M. & Green, D. R. *Nature* **339**, 625–628 (1989).
3. Ucker, D. S., Ashwell, J. D. & Nickas, G. J. *Immun.* **143**, 3461–3469 (1989).
4. Shi, Y. *et al. Science* **257**, 212–214 (1992).
5. Liu, Z.-G. *et al. Nature* **367**, 281–284 (1994).
6. Woronicz, J. D., Calnan, B., Ngo, V. & Winoto, A. *Nature* **367**, 277–281 (1994).
7. Briner, T. J. thesis, MIT (1989).
8. Zacharchuk, C. M. *et al. J. Immun.* **145**, 4037–4045 (1990).
9. Stalder, T., Hahn, S. & Erb, P. J. *Immun.* **152**, 1127–1133 (1994).
10. Ju, S.-T. *et al. Proc. natn. Acad. Sci. U.S.A.* **91**, 4185–4189 (1994).
11. Leo, O. *et al. J. Immun.* **137**, 3874–3870 (1986).
12. Ju, S.-T. *et al. J. Cell. Immun.* **117**, 399–413 (1988).
13. Russell, J. H. *et al. Proc. natn. Acad. Sci. U.S.A.* **90**, 4409–4413 (1993).
14. Russell, J. R. & Wang, R. *Eur. J. Immun.* **23**, 2379–2382 (1993).
15. Kozin, R. L., Babcock, S. K. & Herron, L. R. *J. exp. Med.* **168**, 2221–2229 (1988).
16. Singer, P. A. *et al. J. exp. Med.* **170**, 1869–1877 (1989).
17. Sidman, C. L., Marshall, J. D. & von Boehmer, H. *Eur. J. Immun.* **22**, 499–504 (1992).
18. Zhou, T. *et al. J. Immun.* **147**, 466–474 (1991).
19. Yang, Y., Vacchio, M. S. & Ashwell, J. D. *Proc. natn. Acad. Sci. U.S.A.* **90**, 6170–6174 (1993).
20. Gao, E.-K. *et al. Nature* **336**, 176–179 (1988).
21. Jenkins, M. K., Schwartz, R. H. & Pardoll, D. M. *Science* **241**, 1655–1658 (1988).
22. Sarin, A. *et al. J. Immun.* **153**, 862–872 (1994).
23. Chomczynski, P. & Sacchi, N. *Analyt. Biochem.* **162**, 156–159 (1987).
24. Watanabe-Fukunaga, R. *et al. J. Immun.* **148**, 1274–1279 (1992).
25. Suda, T. *et al. Cell* **75**, 1169–1178 (1993).
26. Sabath, D. E., Broome, H. E. & Prystowsky, M. B. *Gene* **91**, 185–191 (1990).
27. Shen, M. M. & Leder, P. *Proc. natn. Acad. Sci. U.S.A.* **89**, 8240–8244 (1992).
28. Zettlemis, G. *et al. DNA and Cell Biology* **9**, 347–353 (1990).
29. Mohler, K. M. *et al. J. Immun.* **151**, 1548–1561 (1993).
30. Ju, S.-J. *et al. J. Immun.* **144**, 23–31 (1990).

ACKNOWLEDGEMENTS. We thank T. Briner (ImmunLogic), J. Browning (Biogen) and P. Leder (Harvard Medical School) for their help. H. C. is a visiting scientist from the China–Japan Friendship Hospital, Beijing; B.Z.S. is a predoctoral fellow of the Howard Hughes Medical Institute. This study is supported by the NIH, the American Cancer Society, the Arthritis Foundation, the America Institute for Cancer Research and the Lupus Foundation of America, Inc.

Membrane glycoprotein PC-1 and insulin resistance in non-insulin-dependent diabetes mellitus

Betty A. Maddux*, Paolo Sbraccia*, Shinobu Kumakura*, Shlomo Sasson*, Jack Youngren*, Alexander Fisher*, Steven Spencer†, Andrew Grupe†, William Henzel†, Timothy A. Stewart†, Gerald M. Reaven‡ & Ira D. Goldfine*§

* Division of Diabetes and Endocrine Research, Mount Zion Medical Center of the University of California, San Francisco, San Francisco, California 94115, USA

† Genentech Inc., South San Francisco, California 94080, USA

‡ Stanford University, VA Medical Center, Palo Alto, California 94304, USA

MOST patients with non-insulin-dependent diabetes mellitus are resistant to both endogenous and exogenous insulin¹. Insulin resistance precedes the onset of this disease^{2–4}, suggesting that it may be an initial abnormality. Insulin-receptor kinase activity is impaired in muscle, fibroblasts and other tissues of many patients with non-insulin-dependent diabetes mellitus⁵, but abnormalities in the insulin-receptor gene do not appear to be the cause of this decreased kinase activity^{6,7}. Skin fibroblasts from certain insulin-resistant patients contain an inhibitor of insulin-receptor tyrosine kinase^{8,9}. Here we show that this inhibitor is a membrane glycoprotein, termed PC-1 (refs 10, 11). We find that PC-1 activity is increased in fibroblasts from seven of nine patients with typical

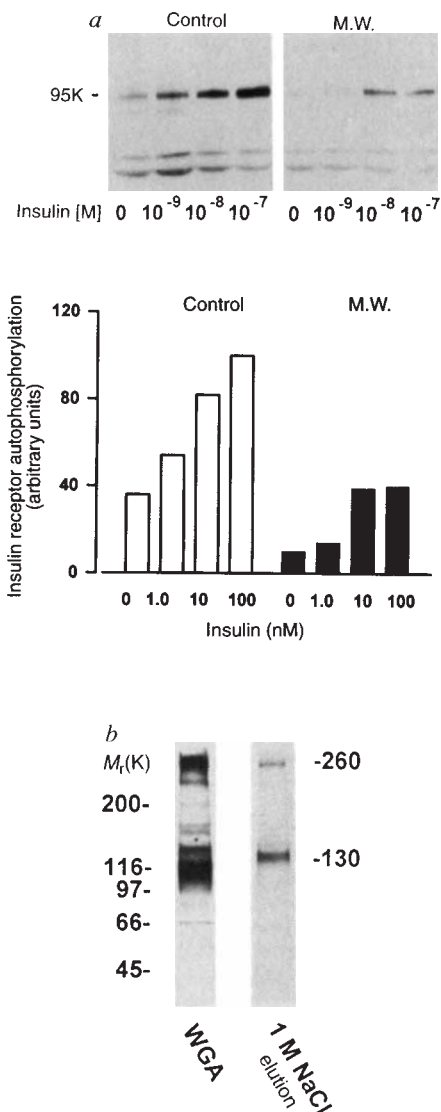


FIG. 1 a, Comparison of insulin receptor autophosphorylation in fibroblasts from patient M.W. and an age- and sex-matched control. A representative autoradiograph of the data (top) and a densitometric scan (bottom) are shown. In the autoradiograph, the location of the 95K insulin-receptor β -subunit is indicated. b, Purification of the insulin-receptor inhibitor on ATP agarose. The 130K and 260K forms of the inhibitor are indicated on a silver-stained gel.

METHODS. a, Cells were grown to confluency and insulin was added for 2 min. The cells were washed three times with PBS at 4 °C, and scraped into 1 ml buffer containing 50 mM HEPES (pH 7.6), 150 mM NaCl, 1 mM PMSF and 2 mM sodium orthovanadate. Cells were pelleted and solubilized in the same buffer without NaCl but containing 1% Triton X-100. Lysates were immunoadsorbed with α -IR3 (a monoclonal antibody specific to the IGF-1 receptor). The depleted lysates were then immunoprecipitated with an anti-insulin-receptor antiserum¹⁹ and immunoprecipitates run on SDS-PAGE in an 8–16% gradient gel (NOVEX) under reducing conditions. Proteins were transferred to nitrocellulose and immunoblotted with an antibody to phosphotyrosine (UBI). b, Fibroblasts (5×10^5) from M.W. were scraped and washed. Cell pellets were solubilized in 50 mM HEPES (pH 7.6), 1 mM PMSF, 2 mM sodium orthovanadate and 1% Triton X-100. Solubilized lysates were applied to an anti-insulin receptor antibody affinity column^{8,9}. The flow-through (depleted of insulin receptors) was then applied to a wheat-germ agglutinin-agarose column (EY Labs). After washing, bound glycoproteins were eluted with 0.3 M N-acetyl-D-glucosamine in 50 mM HEPES (pH 7.6), 150 mM NaCl, 0.01% Tween-20 and 1 mM PMSF. The glycoproteins were desalted using a Centricon 100 filter (Amicon Corp) and applied to a 1 ml ATP-agarose column (Sigma). After washing, proteins were eluted with a NaCl step gradient. Both the WGA eluate and the 1 M NaCl eluate were run on SDS-PAGE and silver stained.

§ To whom correspondence should be addressed.

FIG. 2 a, Western blot analysis of PC-1 protein content in M.W. fibroblasts and controls. The 130K and 260K forms of the protein are indicated. b, Northern blot analysis of PC-1 mRNA content in M.W. fibroblasts and controls. The 8.2- and 3.6-kb species of the insulin receptor mRNA, and the 2.0-kb species of the actin mRNA are shown. METHODS. a, Cells from M.W. and 3 controls were grown to confluency, washed 3 times with PBS and solubilized in 50 mM HEPES (pH 7.6), 1 mM PMSF, 1% Triton X-100 for 1 h at 4 °C. 100 µg protein was then run on SDS-PAGE under reducing conditions, transferred to a nitrocellulose filter and probed with a polyclonal antibody against PC-1 (gift from I. Funakoshi). C₁, human foreskin fibroblast; C₂, 34-year-old female; C₃, 39-year-old male. b, Cells were grown to confluency and poly(A)⁺ mRNA was prepared by a proteinase K digestion method²⁰. 10 µg poly(A)⁺ were subjected to 1% agarose-formaldehyde electrophoresis, transferred to a nitrocellulose filter, and probed with cDNAs to human PC-1 (ref. 10) and β -actin²⁰.

non-insulin-dependent diabetes mellitus. In addition, overexpression of PC-1 in transfected cultured cells reduces insulin-stimulated tyrosine kinase activity. These studies raise the possibility that PC-1 has a role in the insulin resistance of non-insulin-dependent diabetes mellitus.

To characterize and purify the inhibitor, we used dermal fibroblasts from a 42-year-old woman (M.W.) with marked insulin resistance and non-insulin-dependent diabetes mellitus (NIDDM)⁸. Her cells were resistant to insulin, although the insulin-receptor content was normal and purified insulin receptors had normal tyrosine kinase activity^{8,9}. To study the inhibitor in intact fibroblasts, *in vivo* tyrosine kinase activity was assessed by autophosphorylation of the insulin-receptor β -subunit. In control fibroblasts, 1 nM insulin stimulated insulin-receptor β -

subunit autophosphorylation; maximal effects were observed at 10–100 nM insulin (Fig. 1a). The fibroblasts from M.W. required significantly more insulin to stimulate autophosphorylation of the insulin receptor (Fig. 1a).

The inhibitor was purified using a cell-free insulin receptor tyrosine kinase assay^{8,9} to follow the purification process. Solubilized cell lysates were depleted of insulin receptors, and the inhibitor was purified to apparent homogeneity by wheat-germ affinity chromatography followed by ATP-agarose chromatography (Fig. 1b). The inhibitor eluted from the ATP-agarose column as two bands with relative molecular masses of 130K and 260K. The ATP-agarose eluate was >100-fold more potent as an inhibitor of insulin-receptor tyrosine kinase than the wheat-germ eluate. Purification from the cell homogenate was

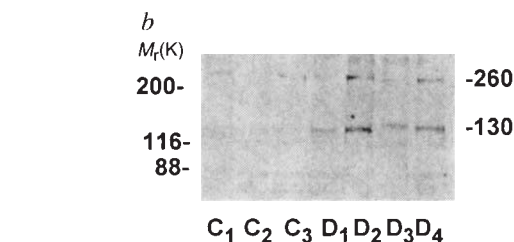
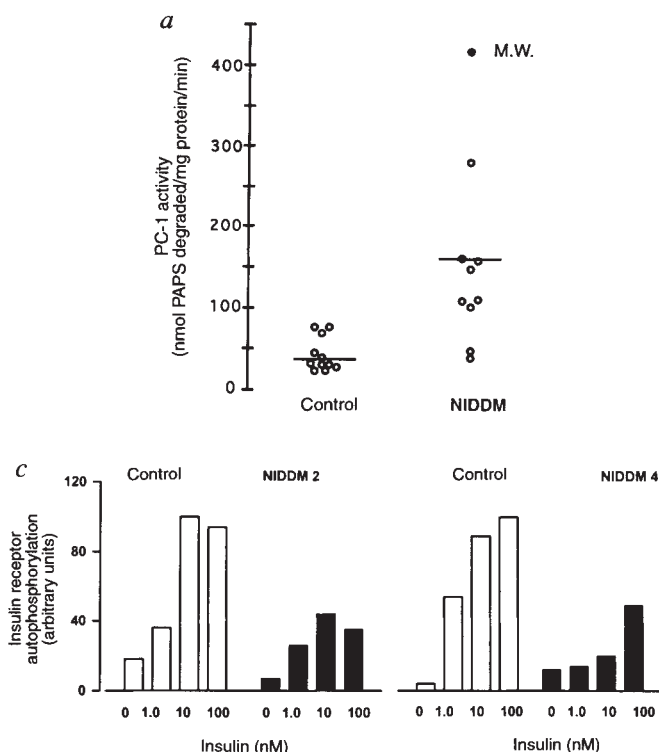
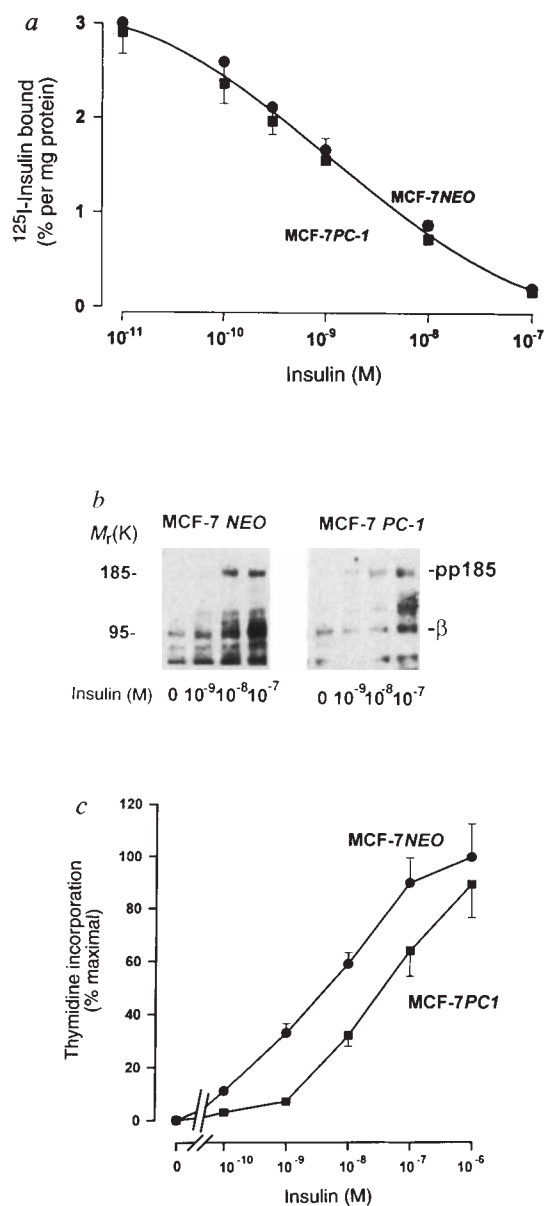


FIG. 3 a, PC-1 activity in dermal fibroblasts from NIDDM patients and controls. b, PC-1 content in dermal fibroblasts from NIDDM patients and controls. The 130K and 260K forms of PC-1 are indicated. c, Insulin-receptor β -subunit autophosphorylation in fibroblasts from 2 NIDDM patients and matched controls. A densitometric scan is shown.

METHODS. a, Fibroblasts from controls and NIDDM patients were grown to confluency, washed as described in Fig. 2, and solubilized in 150 mM NaCl, 1% Triton X-100, 1 mM PMSF and 20 mM imidazole (pH 7.8) for 1 h at 4 °C. Supernatants containing 0.05–3 µg protein were then incubated with 9 nmol [³⁵S]3'-phosphoadenosine, 5'-phosphosulphate (NEN) in the presence of 0.1 µmol MgCl₂ in 20 µl buffer (0.1 M 2-amino-2-methyl-1-propanol-HCl, pH 9.4; Sigma) for 30 min at 37 °C. 25 µl 0.1 M sodium acetate (pH 5.5) was added and samples were boiled for 1 min. Then 0.5 ml of activated charcoal (40 mg charcoal per ml in 20 mM sodium sulphate) was added. After 10 min on ice, tubes were centrifuged, and supernatants counted. b, Fibroblasts from 3 controls (C₁, C₂, C₃) and 4 NIDDM patients (D₁, D₂, D₃, D₄) were analysed for PC-1 content by western blot analysis as described in Fig. 2a. c, Cells were analysed for insulin-receptor β -subunit autophosphorylation as described in Fig. 1a.

FIG. 4 a, Lack of an effect on PC-1 overexpression in MCF-7 cells on insulin binding. A competition-inhibition plot is shown. b, Inhibitory effect of PC-1 overexpression in MCF-7 cells on insulin-stimulated tyrosine-kinase activity. The location of the insulin-receptor β -subunit and IRS-1 (pp 185) are shown. c, Effect of PC-1 overexpression in MCF-7 cells on insulin-stimulated [3 H]thymidine incorporation. In MCF-7 NEO cells, basal incorporation was 19.9 ± 1.8 (mean $\pm$ s.e.m., $n = 4$) pmol per mg protein, and in the presence of $1 \mu\text{M}$ insulin was 36.6 ± 4.7 pmol per mg protein. In MCF-7 PC-1 cells, basal incorporation was 15.0 ± 1.9 pmol per mg protein, and in the presence of $1 \mu\text{M}$ insulin 34.2 ± 4.2 pmol per mg protein. Results are mean $\pm$ s.e.m. of 4 separate experiments.

METHODS. a, Human MCF-7 cells were transfected both with an expression vector²¹ containing the coding sequence of human PC-1, under the control of the cytomegalovirus promoter, and pRK-neo, a selectable marker for neomycin resistance. For controls, MCF-7 cells were transfected with pRK-neo alone. Confluent cells were incubated with 20 pM of [^{125}I]-labelled insulin in the presence or absence of unlabelled insulin¹⁴ for 18 h at 4°C . Binding to plates has been described¹⁴. b, Cells were grown to $\sim 90\%$ confluency, washed, and the medium changed to DMEH-21 with 0.1% BSA. Cells were incubated with insulin for 2 min at 37°C , and then solubilized as in Fig. 1a. Lysates were western blotted as in Fig. 1a but without prior immunoprecipitation. A representative experiment with pRK-neo-transfected and PC-1-transfected MCF-7 cells is shown. c, Cells were prepared as in b, preincubated for 16 h with insulin plus 5 nM $\alpha\text{-IR3}$ (to block the IGF-1 receptor), and incubated for 2 h with $0.5 \mu\text{Ci ml}^{-1}$ [^3H]thymidine. After aspiration and washing, cells were lysed in 0.03% SDS. Lysates were then treated with 10% trichloroacetic acid, and the [^3H]thymidine incorporation measured.



$>1,000$ -fold, as determined by western blotting and enzyme activity.

The eluted 130K band was cleaved with cyanogen bromide and the resulting fragments were separated on a thin (0.5 mm) SDS-polyacrylamide-gradient gel to increase electroblotting transfer efficiency. Bands at 14K and 21K were sequenced (data not shown), and found by Automated Edman degradation to be identical to the corresponding cyanogen bromide peptides of the membrane glycoprotein PC-1 (refs 10, 11).

Western blotting of lysates from M.W.'s cells revealed that there was a 5–10-fold increase in both the 130K (monomer) and 260K (dimer) forms of PC-1 (refs 10, 11) compared with control cells (Fig. 2a). In messenger RNA from M.W., there was a 5–10-fold increase in the principal mRNA species for PC-1 (Fig. 2b); β -actin mRNA was unchanged. As PC-1 has enzyme activity that hydrolyses phosphosulphate bonds, it can be measured using the synthetic substrate 3'-phosphoadenosine-5'-phosphosulphate (PAPS)¹². PAPS hydrolysis in fibroblasts from M.W. was $417 \text{ nmol per mg protein per min}$, a value 10-fold higher than that found for 11 control subjects, 42 ± 7 (mean $\pm$ s.e.m., range 22–76) (Fig. 3a). When extracts of M.W. fibroblasts and purified PC-1 protein were treated with antiserum against PC-1, more than 90% of PC-1 activity and of insulin-receptor tyrosine kinase inhibition was abolished (data not shown).

Fibroblasts were studied from nine additional patients with typical NIDDM (4 males, 5 females; average age, 51 ± 4 years (mean $\pm$ s.e.m.); body mass index, $28.6 \pm 2.7 \text{ kg m}^{-2}$; fasting glucose, $223 \pm 22 \text{ mg dl}^{-1}$; fasting insulin, $28 \pm 9 \mu\text{U ml}^{-1}$) (Fig. 3a). PC-1 activity in these patients' fibroblasts was $128 \pm 25 \text{ nmol per mg protein per min}$ (mean $\pm$ s.e.m.); fibroblasts from two of the nine patients had PC-1 activity within the normal range. In four patients with high PC-1 activity, we found that fibroblast PC-1 content measured by western blotting was increased compared with controls (Fig. 3b). The amount of PC-1 was not increased in the two NIDDM patients with low PC-1 activity (data not shown). In fibroblasts from two of the patients with high PC-1 activity (patients D₂, D₄), insulin-receptor tyrosine-kinase activity was also decreased (Fig. 3c).

Fibroblasts from seven lean, insulin-resistant, non-diabetic subjects were studied; PC-1 from three had increased activity (104 , 109 and $114 \text{ nmol mg}^{-1} \text{ min}^{-1}$). These results indicated that the elevated PC-1 content in fibroblasts from insulin-resistant patients was not the result of *in vivo* hyperglycaemia.

We have tested other human tissues and tissues of an animal model of NIDDM for increased PC-1 (results not shown). We found that in rectus abdominus muscle biopsies from four of six adult human NIDDM patients, PC-1 activity was higher than in controls. In muscle and fat from male Wistar fatty rats, an

animal model of insulin resistance and NIDDM¹³, the number of insulin receptors as well as their tyrosine kinase activity is decreased¹³; we found that the amount of PC-1 in these tissues was raised by 54 and 74%, respectively, compared with controls.

MCF-7 cells, a human breast carcinoma cell line that has been used to investigate insulin action¹⁴, were transfected with an expression plasmid containing PC-1 complementary DNA. Control cells had a PC-1 activity of 10 ± 5 nmol mg⁻¹ min⁻¹, which increased to 405 ± 35 nmol mg⁻¹ min⁻¹ in transfected cells. Overexpression of this protein did not alter insulin-receptor binding (Fig. 4a). With PC-1 overexpression, there was marked inhibition of insulin-receptor tyrosine-kinase activity (Fig. 4b), as measured by both insulin receptor β -subunit autophosphorylation, and phosphorylation of the intracellular protein, insulin receptor substrate-1 (IRS-1)¹⁵. In controls, effects of insulin on both functions were observed at 1 nM and were maximal at 10–100 nM. In transfected cells, effects were detected at 10 nM insulin, and at 100 nM the effect of insulin was one-half that of controls. In MCF-7 cells, insulin stimulates [³H]thymidine incorporation into DNA¹⁴ (Fig. 4c). In control MCF-7 cells, half-maximal stimulation by insulin occurred at concentrations ~8-fold lower than in cells overexpressing PC-1. We have also observed that CHO cells transfected with and overexpressing

PC-1 have decreased insulin-receptor tyrosine-kinase activity (unpublished results).

PC-1 is a class II (cytoplasmic amino terminus) membrane glycoprotein; it is the same protein as liver nucleoside pyrophosphatase/alkaline phosphodiesterase I (ref. 16). PC-1 has been detected in other cells, including placenta, chondrocytes, epididymus, kidney tubules, salivary ducts, brain capillaries, skin fibroblasts, myeloma cells, skeletal muscle and fat (ref. 16, and our unpublished results). The size of PC-1 is 115K–135K, depending on the tissue; PC-1 also exists as a 230K–260K dimer. Human PC-1 is predicted to have 873 amino acids, and maps to chromosome 6q22–6q23 (ref. 11). The extracellular domain of PC-1 cleaves phosphosulphate, pyrophosphate and phosphodiesterase linkages. PC-1 may have threonine-specific protein-kinase activity¹⁷ and is closely associated with the acid fibroblast growth factor receptor¹⁷; it is regulated by transforming growth factor- β (ref. 18).

The function of PC-1 has yet to be identified and so the mechanism by which PC-1 inhibits insulin-receptor tyrosine-kinase activity is only speculation. PC-1 may interact with the insulin receptor to decrease tyrosine-kinase activity, or it may block the insulin-receptor signalling pathway. Further studies will be needed to determine the role of this protein in the insulin resistance of NIDDM. □

Received 1 August; accepted 22 November 1994.

1. Reaven, G. M. *Diabetes* **35**, 1595–1607 (1988).
2. Bogardus, C. et al. *Diabetes* **38**, 1423–1432 (1989).
3. Warram, J. H. et al. *A. Intern. Med.* **113**, 909–915 (1990).
4. Martin, B. C. et al. *Lancet* **340**, 925–929 (1992).
5. Haring, H. & Obermaier-Kusser, B. *Diabetes Metab. Rev.* **5**, 431–441 (1989).
6. Taylor, S. I. *Diabetes* **41**, 1473–1490 (1992).
7. Moller, D. E. & Flier, J. S. *New Engl. J. Med.* **325**, 938–948 (1991).
8. Sbraccia, P. et al. *Diabetes* **40**, 295–299 (1991).
9. Maddux, B. A., Sbraccia, P., Reaven, G. M., Moller, D. E. & Goldfine, I. D. *J. Clin. Endocr. Metab.* **77**, 73–79 (1993).
10. Buckley, M. F., Loveland, K. A., McKinstry, W. J., Garson, O. M. & Goding, J. W. *J. Biol. Chem.* **265**, 29, 17506–17511 (1990).
11. Funakoshi, I. et al. *Archs Biochem. Biophys.* **295**, 180–187 (1992).

12. Yoshida, H., Fukui, S., Funakoshi, I. & Yamashina, I. *J. Biochem.* **93**, 1641–1648 (1983).
13. Greene, S. F., Johnson, P. R., Eifert, K. C., Greenwood, M. R. C. & Stern, J. S. *Obesity Res.* **2**, 432–443 (1994).
14. Milazzo, G. et al. *Cancer Res.* **52**, 3924–3930 (1992).
15. Myers, M. G. & White, M. F. *Diabetes* **42**, 643–650 (1993).
16. Rebbe, N. F., Tong, B. D. & Hickman, S. *Molec. Immun.* **30**, 87–93 (1993).
17. Oda, Y., Kuo, M.-D., Huang, S. S. & Huang, J. S. *J. Biol. Chem.* **266**, 16791–16795 (1991).
18. Huang, R. et al. *J. clin. Invest.* **94**, 560–567 (1994).
19. Forsythe, J., Maddux, B. & Goldfine, I. D. *Diabetes* **35**, 837–846 (1986).
20. Hartmann, K. K. P. et al. *Endocrinology* **127**, 2038–2040 (1990).
21. Suva, L. J. et al. *Gene* **77**, 95–105 (1989).

ACKNOWLEDGEMENTS. We thank I. Yamashina for antiserum to PC-1 and J. W. Goding for human PC-1 cDNA. P.S. was the recipient of an Eli Lilly/EASD Research Fellowship; J.Y. was supported by a National Research Service Award. This work was supported in part by the Walter and Elise Haas Fund.

ERRATA

Evolution of homeotic gene regulation and function in flies and butterflies

Robert W. Warren, Lisa Nagy, Jane Selegue, Julie Gates & Sean Carroll

Nature **372**, 458–461 (1994)

FIGURE 1a of this Letter is shown below; this panel was omitted from page 459. □

a

<i>Drosophila Antp</i>	RKRGRQTYTRYQTL	ELEKEFHFNRL	YLTRRRRIE	AHALCLTERQ	IKIWFQNRRL	KWKKENKTG	EPGSGGEGDE	ITPNSPQ
<i>P. coenia Antp</i>							D.P.N	MS..T..
<i>P. coenia Scr</i>							H.MAS	MNIYPYHmP
<i>P. coenia Ubx</i>							IQAIKEINEQ	QKQQA
<i>P. coenia abd-A</i>							L..LR	AVKEINEQARRER
								EQdrMKqQqekqakleS

Role of the LIM class homeodomain protein Xlim-1 in neural and muscle induction by the Spemann organizer in *Xenopus*

Masanori Taira, Hiroki Otani, Jean-Pierre Saint-Jeannet & Igor B. Dawid

Nature **372**, 677–679 (1994)

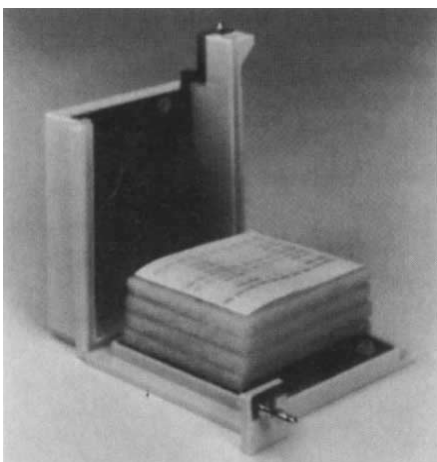
THE panels of Fig. 2b of this Letter should have been labelled: left, FITC; middle, 12/101; right, FITC/TR. □

A buyer's market for protein studies

Check the budget to see if there is room for a MALDI/TOF peptide sequencer, cytoskeletal proteins, a protein purification system, structural clustering software or even an API-electrospray LC/MS system.

THE new Interchange *in vitro* amber suppressor mutagenesis systems from Promega offer methods for **expression of mutagenized proteins** (*Reader Service No. 100*). These systems allow for the study of amino acid replacement effects on protein structure and function *in vivo* and *in vitro*. They are based on the placement of an amber stop codon at a position of interest within a cloned sequence using a site-directed mutagenesis method. System A is specifically designed for the study of amino acid replacement effects on protein folding and packing. System B is for the study of amino acid replacement effects on catalytic activity, DNA binding and kinase activity.

Novex's MultiMark standard is designed to give sharp, coloured molecular weight bands during electrophoresis runs and tight distinct bands after transfer (*Reader Service No. 101*). Dyes are covalently bound to the **standard proteins** and will not be removed by usual staining and destaining procedures. Novex has controlled the dye/protein binding reaction to minimize lot-to-lot variation with respect



Multicolour molecular weight standards.

to migration and band sharpness. The MultiMark pre-stained standard should save time as it requires no mixing, reducing, weighing or heating. The standard includes nine proteins ranging from 4 to 205 kilodaltons and is said to perform equally well in both the Tris-Glycine and Tris-Tricine buffer systems. It calibrates straight percentage gels from 4 to 20 per cent and is said to be suitable for gradient gels.

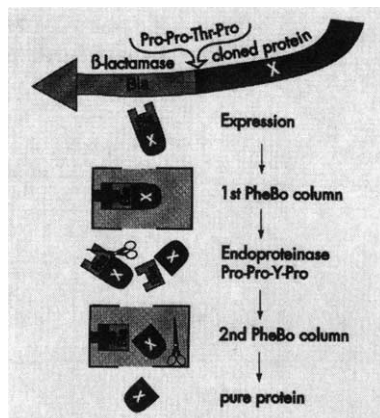
Oxford GlycoSystems has developed a range of **glycoamino acids** which are supplied as stable, dry powders (*Reader*

Service No. 102). A base labile fluoroenyl-methyloxycarbonyl group protects N α groups while side-chain monosaccharide hydroxyl residues are protected by O-acetylation. N-glycosyl (Asn) and O-glycosyl (Ser/Thr) derivatives are soluble in most of the organic solvents used in solid-phase peptide synthesis and the protecting groups used are stable to repetitive coupling and deprotection steps. These glyco-amino acids are part of a wider range designed for glycopeptide and glycoprotein remodelling.

Cytoskeleton specializes in the manufacture of **cytoskeletal proteins**. New additions to their line include tubulins, actins, molecular motor proteins such as kinesin, microtubule and actin binding proteins such as MAP2 and gelsolin (*Reader Service No. 103*). Several kits are also available to aid the researcher in studying the cytoskeleton and its functions. These include microtubule motor protein motility assays and a microtubule binding protein identification kit. All proteins are quality controlled for high purity and biological activity — kinesin motor protein, for example, is tested for purity, microtubule stimulated ATPase activity and its ability to support motility.

■ Protein engineering

MoBiTec now offers a new system for **fusion protein cloning and purification** (*Reader Service No. 104*). The PheBo-System clones the protein of interest next to the leader protein, β -lactamase. The expressed fusion protein remains in the cytoplasm (vectors pHKcyto1 and pHKcyto2) or is released into the periplasm (vectors pHKperi1 and pHKperi2). β -lactamase allows efficient protein purification on reusable PheBo-affinity columns. The leader protein is cut



The PheBo-System from MoBiTec.

off by the site-specific endoprotease Pro-Pro-Y-Pro, which also cleaves inclusion bodies. A second passage through the same column yields the pure protein of interest, and β -lactamase and the endoprotease remain on the column.

PharMingen has released the Baculovirus GST **fusion vectors** and kits (*Reader Service No. 105*). The kits contain all necessary reagents and technical information for GST fusion protein expression and purification. These recombinants are expressed in insect cells and are said to be mainly soluble and functionally active. The GST tag allows purification of recombinant products to be performed in a single, non-denaturing affinity chromatography step using glutathione-coupled beads. The tags can be removed with thrombin in the case of pAcG2T and blood coagulation factor Xa in the case of pAcG3X.

■ Structural software

Chemical Design offers a new module for Chem-X software — Fast Clustering can **group databases** in excess of 50,000 structures into subsets of similar compounds (*Reader Service No. 106*). Representative members of these subsets are then extracted to provide a further subset reflecting the diversity of the database screening. The algorithm used in Fast Clustering groups molecules in the three-dimensional database according to the similarity of their pharmacophores. The company tests show that databases in excess of 40,000 structures can be clustered in under 13 h on a UNIX workstation. This software is supported on a range of hardware platforms including MS-Windows, DOS, Apple Macintosh and UNIX workstations such as Silicon Graphics, IBM RS/6000 and Digital Alpha AXP.

SciProtein from SciVision is designed as a tool to help bridge the **software gap** between primary sequences and existing three-dimensional structures (*Reader Service No. 107*). With this program users can explore homology, motif identity, sequence property maps and construction of secondary structures. The search-and-find feature allows users to locate files from a hard drive or PDB CD-ROM by name prefix or extension. Selected files may then be loaded directly from databases. Amino acid property maps present information on hydrophathy, antigenicity, mutability, solvent accessibility, and

residue flexibility. SciProtein is able to select residue averaging, filter size and export values to Excel. There is also a three-dimensional building feature in HyperChem, by segment or total sequence.

For identification of lead compounds using computer-aided drug design MDL Information Systems offers the **integrated scientific information system, ISIS** (Reader Service No. 108). MDL's database searching tools and three-dimensional structural databases are designed for developing and validating pharmacophore hypotheses that are selective for bioactive molecules. Macromolecular studies that lead to a small molecule hypothesis can be input through ISIS/draw. Then with three-dimensional searching a selection of commercially available chemicals or other compounds can be imported to the workstation. Structure-based drug design with ISIS can be accomplished through a single, consistent interface that conforms to popular operating systems, including: Apple Macintosh, SGI IRIS and IBM PC running MS-Windows.

A new data sheet from Perkin-Elmer describes the company's **capillary electrophoresis bibliography** (Reader Service No. 109). This database contains nearly 1,400 published citations from 1984 to 1993. Each citation includes information on the author, title, journal, year published and key words. The database allows flexible searching by author, title and key words. The database was developed for use with both Apple Macintosh and IBM or compatible PCs. Each bibliography package consists of a data disk and a demonstration version of the software, which allows easy access to the database. A free copy of the data sheet on the bibliography database is available.

■ Protein separations

Owl Scientific has introduced the new Bandit tank **electroblotter** for transfer of high molecular weight proteins (Reader Service No. 110). This system will transfer two gels at a time with an area of 20×18 cm per gel. Bandit will transfer proteins up to 200 kilodaltons in times stated as less than 60 min at 2 A constant current with cooling. An optional cooling module is available to maintain low buffer temperature and speed transfer. This electroblotter requires 3 litres of buffer, incorporates solid-plate electrodes for long-term durability and includes two colour-coded transfer cassettes for error-free transfer. It also has an area at the base of the tank reserved for a stir bar to prevent ion depletion and temperature variation within the tank.

Pharmacia Biotech has introduced two **silver staining reagent** kits that are said to bring speed, convenience and high sensi-

tivity to the staining of proteins and nucleic acids in polyacrylamide electrophoresis gels (Reader Service No. 111). The silver staining kit for proteins and the silver staining kit for DNA each contain ready-to-use solutions and pre-weighed chemicals to reduce preparation time and increase reliability. The kits give colourless backgrounds in most gel electrophoresis systems and are for use with pre-cast gels as well as with thicker, minislabs polyacrylamide gels. The silver staining kit for proteins gives results in about 2 h and detects most proteins in the nanogram range — stated to be about 100 times more sensitive than Coomassie Brilliant Blue.

The new BioLogic **protein purification system** from Bio-Rad is intended for analytical and preparative-scale separations (Reader Service No. 112). This modular system can be used in cold-rooms and is designed to be flexible to meet individual user needs. The controller provides single-point control of all system components through Windows-based software for programming, set up, monitoring and documentation. The BioLogic workstation



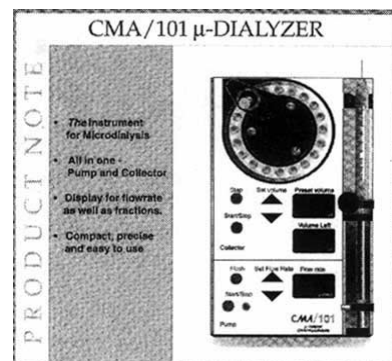
The BioLogic protein purification system.

contains dual-position pumps with high-pressure dynamic mixing for accurate, reproducible gradients. Control circuitry for ultraviolet detection, conductivity detection and value control are all based in the workstation. The system operates low- and high-pressure (up to 6,900 kPa) columns expanding protein purification capabilities to include ion-exchange, affinity, hydroxyapatite, HIC, size-exclusion and reverse-phase columns.

For a new **protein staining** option, B/T Scientific Technologies has released Blv stain (Reader Service No. 113). The company states that destaining a Blv gel requires less than 10 min and that it has a two-fold increase in sensitivity over Coomassie Blue with no loss of linearity — the intensity of the band is linear with protein concentrations in the nanogram-to-microgram range. Blv does not rely on caustic chemicals and is said to reduce solvent waste by about half.

CMA/Microdialysis has launched the CMA/101 μ -Dialyzer — an integrated unit consisting of a syringe pump and a

fraction collector (Reader Service No. 114). The syringe size, flow rate, flush feed and fraction size are easily set and displayed on the user panel. The dialyzer is precalibrated for 1- or 2.5- ml syringes and



CMA/101 μ -Dialyzer.

flow rates between 0.1 and $20 \mu\text{l min}^{-1}$. Up to 20 samples of $1 \mu\text{l}$ to $50 \mu\text{l}$ can be collected in plastic or glass vials. Its small size enables it to be situated close to the experimental set-up.

■ Catalogues

Amersham Life Science has produced a 36-page catalogue specifically highlighting its range of products, both radioactive and non-radioactive, for the **labelling and detection of peptides** (Reader Service No. 115). The catalogue features the main techniques of protein research and is divided into sections covering: western blotting, total protein detection and immunochemistry; cellular protein labelling; post-translational modification; and *in vitro* translation systems. Included in each section are full details of all products, background information and typical results illustrating how particular products can be used in practice. A fifth section provides information on related products such as molecular weight markers and membranes. The catalogue is available free upon request.

New from Hoefer is the company's **protein electrophoresis application guide** (Reader Service No. 116). The guide was developed to accompany the firm's electrophoresis workshops. It has complete illustrations and detailed procedures for conducting one- and two-dimensional polyacrylamide gel electrophoresis, flat-bed isoelectric focusing and immunoblotting. Additional information is included on staining, membrane processing, gel drying and molecular weight determination.

■ Reagents

Merck has introduced Peflabloc SC **serine-protease inhibitor** to help protect against loss of activity and reduced yields (Reader Service No. 117). The inhibitor is said to be an irreversible inhibitor of all serine proteases and to exert its effect even at low concentrations. The compound is usually used in concentrations of $50 \mu\text{M}$ to 200

μM , but a final concentration of 1 mM to 5 mM is recommended for complete inhibition of all serine proteases. Peplabloc SC is stated to be readily soluble in aqueous or buffer solutions and to be stable at a number of temperatures and pHs. In terms of toxicity, it is stated as being 354-times less toxic than PMSF (phenylmethyl sulphonyl fluoride) and DFP (diisopropyl fluorophosphate).

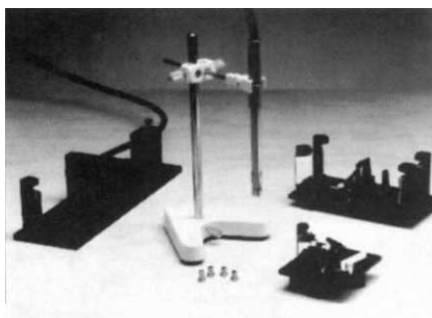
Calbiochem-Novabiochem has released a new non-ionic detergent similar to octyl-glucopyranoside and HECAMEG but with a higher stated CMC for **enhanced protein solubilization** (Reader Service No. 118). NOGA (n-octanoyl- β -D-glucosylamine) has a stated CMC of 80 mM and a low ultraviolet absorbance above 250 nm. This non-ionic detergent can be removed by dialysis and is gentle to membrane proteins.

■ Instruments

Hewlett-Packard has introduced their atmospheric-pressure-ionization electrospray liquid chromatography/mass spectrometer (API-electrospray LC/MS) system (Reader Service No. 119). This system offers Windows-based software for instrument control and data interpretation. Trifluoroacetic acid (TFA) sensitivity-enhancement chemistry provides sensitivity even in the presence of strong ion-pairing agents such as TFA. The online characterization of peptide maps at picomole levels helps to achieve faster regulatory approvals, shorten product-development cycles and to reduce risk of product recall. The API-electrospray LC/MS system is also intended for laboratories that analyse proteins and peptides in the picomole-to-femtomole range for their molecular weight. Molecular weight determinations for peptides, intact proteins and oligosaccharides up to 150,000 daltons are stated as being accurate to better than 0.02 per cent.

For sensitive, accurate **peptide sequencing** Finnigan Mat offers the Lasermat 2000 in conjunction with a novel volatile degradation reagent (Reader Service No. 120). The instrument uses a new process that is designed to permit rapid

processing of samples in parallel. Matrix-assisted laser desorption/ionization time-of-flight spectrometers enable rapid analysis of peptide mixtures at high sensitivity. Recent ladder sequencing techniques offer sub-picomole detection limits and the potential to process multiple samples in parallel. A nested set of peptides is generated by adding equal aliquots of starting peptide to each cycle and driving both coupling and cleavage reactions to completion. All excess reagent, buffer and reaction by-products are volatile and are removed under vacuum, eliminating extractive and transfer loss of peptides. Trifluoroethylisothiocyanate is used to meet the specific requirements of this system. The mass accuracy of Lasermat is stated as ± 0.5 daltons. With sample loads as low as 50 femtomoles, the presence of residual buffer salts and excess reagent are said to have no observable effect on the spectrum generated. Samples can be run unattended using a 48-position autosampler with OPAL software which optimizes laser power and aim.



ATI Unicam remote spectrometry probe.

New from ATI Unicam is a range of **accessories for ultraviolet and visible light spectrometers** (Reader Service No. 121). The system includes an ultraviolet/visible fibre probe, an integrating sphere and a specular reflectance accessory. The fibre probe can be used for *in situ* analysis of process streams and in areas where it is difficult or undesirable to choose conventional sampling techniques.

■ Labware

For pyrogen-free water there is the **EASYpure compact deionization system** (Reader Service No. 122). The UF system, from Barnstead/Thermolyne, is designed to provide pyrogen-free water for volumes of less than 15 litre per day. The 10,000 molecular weight ultrafilter provides less than 0.005 EU ml^{-1} pyrogen levels and greater than four-log reduction of pyrogens with flow rates of up to 1.5 l min^{-1} . Water is filtered at the point of use with 0.2 μm absolute filter.

Life Sciences International extends its range of power sup-

plies for electrophoresis with the EC-Apparatus Series-90 range (Reader Service No. 123). The EC6000-90 provides a power output of 6000 V, 200 mA and 200 W and is designed to accommodate most electrophoretic techniques, including ultra-thin and long DNA sequencing gels, SDS-PAGE and IEF. This unit is designed to fit in crowded bench areas, to be portable and can be stacked. There are four output connectors that enable the simultaneous running of multiple gels. The power supply provides an easy-to-read display of operating parameters and is fitted with a 100-h timer to accommodate preprogrammed runs. The automatic restart function prevents the loss of samples in the event of a temporary power failure. □

These notes are compiled by Brendan Horton from information provided by the manufacturers. For more details, fill in the reader service card bound inside the journal.

Erratum

PolyAtract 9600 series mRNA isolation system, as evaluated 26 January product review (Nature 373, 367; 1995), is a Promega product. The company name was omitted.

ADVERTISEMENTS

PEPTIDE and DNA

Peptide and DNA engineering specialist:
Phosphoryl, Thiophosphoryl, Biotinyl,
Chloromethyl, Cyclic, Dimeric, Partial protected,
FITC, DNP, pNA, Phosphothionate and more ...

IonSpray MS and LC/MS

Synthetic or natural compounds analysis:
Organic, inorganic, or coordination compounds.
Drug metabolites, Enzyme digests, etc.
\$30/sample (MS), \$250/sample (LC/MS)

Call for more information (800-597-7873) or
Fax us sequence for quotation (510-371-1156).
PeptidoGenic Research & Co. Livermore, CA

Reader Service No.3

OLIGONUCLEOTIDES

Synthesised, purified and shipped in 3 days

£2.50 per base

(All inclusive)

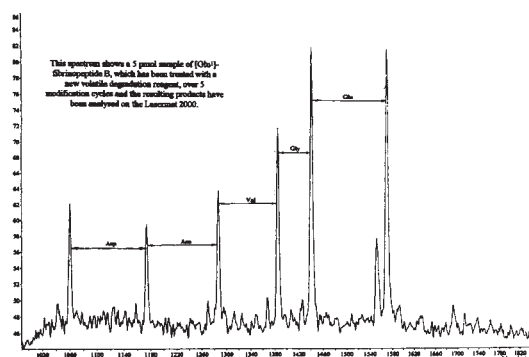
0.2 μM scale (approx. 500 μg)

- Antisense (completely or partially sulphurised)
- Extensive range of labelling and modifications
- Full design and analysis facilities
- Gene cassettes, bulk order and contract discounts

Molecular Medicine Unit

King's College School of Medicine & Dentistry
Tel: 071 346 3126 Fax: 071-733 3877

Reader Service No.10



Fibrinopeptide B spectrum on Lasermat 2000.

Nature's sister journals in review

As a service to those readers of *Nature* who do not see its sister journals, the following pages refer to some of the content of *Nature Genetics*, *Nature Structural Biology* and *Nature Medicine*.

Where history meets genetics: the Ashkenazi population

THE suggestion that many of the genetic diseases common in the Ashkenazi Jewish population originated as recently as the fifteenth century has been put forward by a group of nine US geneticists writing in the current issue of *Nature Genetics* (N. Risch *et al.*, *Nature Genet.* 9, 152–159; 1995).

On the basis of genetic linkage studies of the inheritance of the disease called idiopathic torsion dystonia (ITD), the group also argues that the origin of these diseases may have been in the northern part of the region of north Central Europe known as the Pale (including Lithuania, the Ukraine and parts of Poland) to which Russian Jews were banished by the czars of the time.

Part of the general interest of the study is that it points to the opportunities (and the difficulties) of inferring the history of a population from its genetic structure.

ITD is an inheritable condition marked by twisting movements or postures affecting particular groups of muscles. Although the condition is found in non-Jewish populations, it appears to be most frequent in the Ashkenazim. The group

now reporting has previously localized the gene, *DYT1*, on chromosome 9 (9q34) in which a mutation appears to be responsible for the early-onset version of the disease.

The new development rests on the use of recently discovered genetic markers in the neighbourhood of *DYT1*, and on the recruitment to what is plainly a continuing study of new families of Ashkenazi descent. One outcome of the linkage studies is that ITD is not recessively inherited,

as previously thought, but rather is dominantly inherited, but with low penetrance; only about a third

of those carrying one copy of the gene, and perhaps as few as 10 per cent, develop overt symptoms.

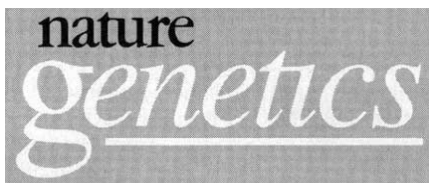
The estimation of the antiquity of the mutation responsible for ITD rests on the frequencies of genetic linkage between the mutation and the nearby genetic markers, which are uniformly greater than the values expected if genetic recombination were long-standing and which allow the authors to calculate the number of

generations for which genetic recombination (at meiosis) has been at work since the first appearance of the mutation.

The results of these calculations are small numbers. The best estimate is that 12 or 13 generations have passed since the appearance of the mutation, but the allowable range is from 8 to 22 generations. At 25 years for a generation, the origin of the mutation is within the range 1400 to 1750 AD, and is most probably dated at 1650 AD (which corresponds to the epoch of the expulsion of Russian Jews to the Pale).

One puzzling feature of the authors' analysis is their attempt to compare the present high frequency of the gene, between 0.017 and 0.050 per cent, with its likely frequency in the early stages of its history. In the absence of accurate data on the mediaeval population of the Pale, the frequency of the *DYT1* gene then can be estimated only inaccurately, but is reckoned to have been less than at present, perhaps by a factor of 10.

How could a dominant disease gene, on which natural selection would have worked directly, have become more and not less frequent over the generations? The authors remark that the Jewish population of north Central Europe has fluctuated markedly as pogroms have been interspersed with periods of very rapid population growth, and that the increase of the frequency of the *DYT1* gene may have segregated by chance in the richer and thus more productive families.



Nature Genetics — other points

Three mutations of the gene coding for the α -tocopherol transfer protein (α -TTP) on chromosome 8 (8q13) have been implicated as responsible for the majority of cases of inherited ataxia with vitamin-E deficiency by an international collaboration led by Professor Michel Koenig at Strasbourg; the findings substantiate the value of vitamin E as a prophylactic of progressive neurological damage.

A technique for identifying abnormalities of chromosomes near their ends (telomeres) has been used to identify 3 patients out of 99 with severe mental retardation in whom there are either deletions or rearrangements of sub-

stantial amounts of genetic material, confirming the suspicion that abnormal subtelomeric regions are particularly fragile, and with mental retardation as a consequence.

The longest-known human gene, that for the protein dystrophin which consists of a stretch of DNA 2.34 million base-pairs long comprising 79 functional elements (exons) whose mutations are involved in some forms of muscular dystrophy, requires 16 hours for complete transcription in cultures of muscle cells, but the unwanted parts (the introns) of the transcribed RNA molecule are excised as the transcript forms. □

Head injury and Alzheimer's

The supposed genetic predisposition to late-onset Alzheimer's disease (AD) has been confirmed in what may seem a bizarre study — the post-mortem examination of the brains of people dying after head injuries. The study is described in this month's *Nature Medicine* by James A. R. Nicoll and David I. Graham from the Southern General Hospital in Glasgow, Scotland, and Gareth W. Roberts from SmithKline Beecham Pharmaceuticals.

The genetic link with AD involves the gene responsible for the structure of the lipoprotein molecule ApoE. The protein occurs in three distinct isoforms, each derived from one of three allelic versions of the gene known as E2, E3 and E4. Indi-

viduals homozygous for the last of these make up a large proportion of those who the last of which is associated with the occurrence of late-onset AD.

The role of the E4 version of the lipoprotein in the formation of the extracellular plaques in brain tissue characteristic of AD remains controversial; some hold that the isoform interacts directly with the protein β -amyloid to form the materials of the plaque, others that the interaction is indirect.

The new study has been stimulated by

nature medicine

earlier observations that head injury is a risk factor for AD and that many boxers develop a form of dementia similar to AD. The Glasgow group used brains from 90 people with head injury (mostly in road accidents) who survived for less than two weeks, searching for the β -amyloid plaques characteristic of AD and correlating that with the genetic status of the ApoE production.

In total, 23 of the whole group were found to carry β -amyloid plaques, but the six who were homozygous for the E4 allele all carried these tell-tale signs. By contrast, of the 30 people with just one E4 allele, only 11 had developed plaques by the time of death.

The inference from these observations is that head injury can trigger the formation of extracellular plaques in people who are susceptible to AD. (It is known from *in vitro* studies that increased production of amyloid β -protein may be among the immediate responses to trauma of brain tissue.) But to the extent that

the formation of AD-like plaques appears within two weeks or less of head injury, the observations may also point to the mechanism whereby trauma (as in boxing) causes long-term injury.

More Alzheimer's

The same issue of *Nature Medicine* includes a demonstration in *in vitro* studies that analogues of the amyloid β -protein forming plaques in Alzheimer's disease stimulate the activity of tissue-type plasminogen activator (TPA) and that the effect is enhanced when the β -protein molecules are coagulated; this observation, suggesting a means by which blood vessels may be damaged, may be related to the causation of hereditary cerebral haemorrhage and even the increased incidence of brain haemorrhage in patients being treated with TPA after a heart-attack.

In the same issue, a group based at Toronto and at Kingston in Ontario describes studies with mice and *in vitro* showing that sulphonated and sulphated hydrocarbon molecules of small molecular weight can inhibit the formation of amyloid plaques of all kinds, not only those occurring in Alzheimer's disease; an accompanying comment notes that the toxicity of these compounds has yet to be determined.

Sense and serpin mutations

A number of human disease conditions, such as emphysema, thrombosis and angioedema, are the result of mutations in serine-proteinase inhibitors (serpins) naturally present in the blood. A group working in the Department of Haematol-

nature structural biology

ogy at the University of Cambridge has now rationalized the effects of these mutations based upon the recently published structures of members of the serpin protein family (P. E. Stein & R. W. Carrell *Nature struct. Biol.* 2, 96; 1995).

Serpin molecules consist of around 400 amino acids folded into a characteristic three-dimensional shape consisting of three β -sheets and nine α -helices. Although the serpin family includes some proteins lacking activity as proteinase inhibitors (and which may, for example, function as hormone transporters instead), the function of all the inhibitory

Nature Structural Biology — other points

The isomer of human insulin formed by making an unnatural cysteine bridge within the A chain is trapped in a metastable kinetic state whose study (by NMR) throws light on the protein-folding potential of the native hormone.

Intermediate structures in the refolding of denatured membrane protein bacteriorhodopsin have been provisionally identified by fluorescence spectroscopy, suggesting an ordered sequence in the natural formation of the seven transmembrane protein elements.

A structure of the integral membrane protein phospholamban, a calcium channel involved in the regulation of calcium concentration in heart-muscle cells, has been derived by computation; the predicted structure includes a channel lined with hydrophobic residues and with five-fold symmetry. □

serpins depends on the insertion of an exposed loop into the active site of the target proteinase molecule, thus mimicking the specific substrate of the enzyme.

Stein and Carrell remark that the serpin loop appears to be under considerable tension; when cleaved, either slowly by the serpin's target proteinase or by other proteinases, a dramatic conformational change separates the two halves by almost the entire diameter of the protein, with half of the loop inserting into the middle of one of the serpin's β -sheets.

But even when the loop remains intact, it may adopt a variety of conformations. In its active form, the uncleaved loop can be partially inserted into the β -sheet, while full insertion of the loop yields an inactive or latent conformation.

This leads the authors to an analysis of close on 100 mutations of serpin molecules that are associated with disease, and to a demonstration of how the dysfunction of the molecules can be interpreted in terms of structure and conformational flexibility. For example, mutations that increase the size of amino acids at one end of the reactive-centre loop prevent pre-insertion and thus make inhibition less efficient.

Apart from the intrinsic interest of the data, this monumental analysis may be regarded as an illustration of how the results of high-resolution structural studies can now be combined with clinical observations of diseased patients to yield a coherent view of the functioning of crucial processes in molecular physiology. □

Nature Medicine — other points

The dormancy of metastases associated with primary tumours (which accounts for the difficulty of their early detection), which can be induced in model tissues by blocking the development of blood vessels by antibodies against angiogenic factors, may be a dynamic process in which external inhibition induces increased apoptosis (programmed cell death).

It may be possible to reduce the severity of gastrointestinal side effects of non-steroid anti-inflammatory drugs (of which aspirin is the most commonly used) by complexing them with phospholipid molecules comparable with those found naturally in the stomach wall. □